ORGANISMAL PERFORMANCE ASSAYS: AN INNOVATION IN PHARMACEUTICAL SAFETY ASSESSMENT

by

Shannon Marie Gaukler

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

The University of Utah

August 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Shannon Marie Gaukler**

has been approved by the following supervisory committee members:

Wayne K. Potts	, Chair	**05/01/2014** Date Approved
David R. Carrier	, Member	**05/01/2014** Date Approved
M. Denise Dearing	, Member	**05/01/2014** Date Approved
J. Greg Lamb	, Member	**05/01/2014** Date Approved
J. Michael McIntosh	, Member	**05/01/2014** Date Approved

and by **Neil Vickers** , Chair/Dean of

the Department/College/School of **Biology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Too often, prescription medications once thought safe for human use cause unacceptable adverse health consequences and fail during clinical testing or are recalled after public release. One major contributing factor to this problem is the inability of current *in vitro* and *in vivo* toxicity assessment methods to adequately detect toxicity caused by pharmaceuticals during preclinical trials. We have developed a novel toxicity discovery and quantification method known as the organismal performance assay (OPA) that is capable of revealing mammalian health consequences with great sensitivity. OPAs utilize genetically diverse wild mice (*Mus musculus*) that compete amongst each other for limited resources in seminatural enclosures. OPAs achieve their sensitivity because they are unique in that no hypotheses pertaining to specific mechanisms of action are needed. Rather, whole organismal performance in its entirety is measured by OPAs in a seminatural environment. As a result, any substance or treatment that reduces performance of almost any physiological system is likely to be detected by OPAs as the endpoint measures (reproductive success, survival, and male competitive ability) of performance are integrated across physiological systems. In order for individuals to become successful in OPAs, high performance of most physiological systems is needed. To demonstrate OPAs sensitivity and the potential utilization during preclinical trials, we assessed three pharmaceuticals: two of which have already been recalled from the market (cerivastatin and rofecoxib) and one that is still available, but is suspected of causing

congenital malformations (paroxetine). First, we examined organismal performance consequences associated with exposure to cerivastatin and observed that exposed females had reduced reproduction and males had reduced weight, competitive ability, reproduction, and survival. Next, we tested the safety of paroxetine exposure and found that females had reduced reproduction but appeared to recover from fitness declines after prolonged periods without exposure; however, exposed males suffered lifetime consequences of reduced weight, competitive ability, reproduction, and survival. Then, we assessed the safety of rofecoxib and found that exposed females had an increase in reproduction but the exposure did not affect male weight, competitive ability, reproduction, or mortality. Lastly, as we observed that paroxetine-exposed breeders produced litters with a significant sex-ratio bias towards female offspring, we then analyzed these data in the context of both the Trivers-Willard hypothesis and the cost of reproduction hypothesis. We found support for the Trivers-Willard hypothesis as females from biased litters out-reproduced their brothers in seminatural enclosures. These data suggest that OPAs are more sensitive in detecting physiological adversity than other currently used assays and by implementing OPAs into preclinical trials, they could potentially improve the quality of pharmaceutical safety assessment.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ACKNOWLEDGMENTS

I would first like to extend my sincerest appreciation and gratitude to my advisor, Wayne Potts, for providing me with the opportunity to study in his lab and for his mentorship over the past five years. I would also like to thank my committee, Drs. Dave Carrier, Denise Dearing, Greg Lamb, and Michael McIntosh, for their constructive criticism and feedback. I offer a special thanks to the Potts' Lab members: Linda Morrison, for providing encouragement and advice; James Ruff, for statistical advice and proof reading countless documents; Doug Cornwall, for his assistance with R code, and to Earl Middlebrook for feedback. I extend my appreciation to all of the undergraduate students that contributed significantly, for without them, I would still be genotyping mouse populations. These students include Michelle Bartlett, Stephanie Eddy, Tessa Galland, Kirstie Kandaris, David Kircher, Nicole Liu, Elyse Schwab, Mirtha Sosa, Tristan Underwood, Chloe Young, and Elizabeth Young. I would also like to thank Ahmed Bwika and Jon Gale for maintaining our mouse colony, and Shannon Nielsen and Renae Curtz for all of their assistance with the technical aspects of graduate school. Finally, I would like to thank my family and friends for their continued support and a special thanks to my husband, Ben Sutter, for all of his encouragement, a tremendous amount of support, and for providing a distraction from science. This dissertation was funded by the University of Utah's Technology Commercialization Program.

CHAPTER 1

USING ORGANISMAL PERFORMANCE ASSAYS TO ASSESS THE SAFETY OF PHARMACEUTICALS

1.1 Abstract

Too often, prescription medications once thought safe for human use are found to cause adverse health consequences and fail during clinical testing or are recalled after public release. One major problem is the inability of current toxicity assessment methods to detect many toxicities caused by pharmaceuticals during preclinical trials. Thus, there is a great need for a broad, sensitive, and functional assay to test the safety of pharmaceuticals. We have developed an alternative research method that is capable of revealing mammalian health consequences, known as the organismal performance assay (OPA). OPAs utilize genetically diverse wild mice (*Mus musculus*) where individuals from different treatments compete amongst each other for limited resources in seminatural enclosures. Organismal performance is measured in terms of Darwinian fitness (i.e., reproduction) and factors that contribute to fitness (e.g., survival and social dominance). OPAs are sensitive, broad, and functional because they demand high performance of most physiological systems simultaneously in order for individuals to be successful. OPAs have proven powerful in detecting adverse health effects from genetic inbreeding, bearing a selfish gene, and consuming added sugar at human relevant levels

that were all missed by conventional methodologies. The implementation of OPAs during preclinical trials has the potential to dramatically increase the detection rate of harmful substances before they move onto human testing.

1.2 Background

The drug approval process is costly and requires several years of research. From the time a new chemical entity (NCE) is discovered until market launch, it is estimated to cost more than $1.4 billion and typically takes 12-15 years of preclinical and clinical research (Miller 2012). Despite this large investment, 73% of pharmaceuticals that pass preclinical trials fail during clinical testing (Lipsky and Sharp 2001) and 10% of all Food and Drug Administration (FDA)-approved pharmaceuticals fail after market launch due to unacceptable adverse drug reactions (ADRs) (Schuster et al. 2005). These harmful drugs are then recalled from the market or labeled with black box warnings that warn consumers of the serious risks (Lipsky and Sharp 2001, Schuster et al. 2005). For example, troglitazone (Rezulin®, Parke Davis/Warner-Lambert) was a prescription anti-diabetic drug, which was discovered to cause liver toxicity and was linked with 83 cases of liver failure. Troglitazone was on the market for three years before the FDA recommended the drug be recalled (Faich and Moseley 2001). However, unlike troglitazone, some ADRs are not discovered until several years later; for example, fenluaramine (Pondimin®, Wyeth-Ayerst Laboratories) and phentermine (Lonamin®, Wyeth-Ayerst Laboratories), or Fen-Phen, was on the market for 24 years before the link to valvular heart disease was discovered (Connolly et al. 1997). With 10% of pharmaceuticals failing after approval, it is clear that one of the biggest challenges

pharmaceutical companies face during the approval process is the identification of unpredictable harmful physiological effects.

Preclinical trials are the initial phase of drug development and include both *in vivo* and *in vitro* safety testing. One of the main objectives of *in vitro* testing is to assess drug-induced toxicities by tissue culture or organ specific assays. As cardiac and liver toxicity are the two primary causes of drug failure, they are routinely addressed in preclinical trials (Schuster et al. 2005). The hERG (human EtheR-a-Go-go) channel is a potassium channel that is important to the electrophysiology of the heart. Several drugs have inadvertently inhibited the function of the hERG channel and consequently, this channel has been evaluated during drug development (Brimecombe et al. 2009). *In vitro* assays used to assess hepatotoxicity include the use of both cell culture and liver slices (Dambach et al. 2005). Liver enzyme levels such as alanine aminotransferase (ALT), aspartate aminotransferase (AST), and lactate dehydrogenase (LDH) are monitored and the induction of cytochrome P450 (CYP450) is tested for abnormal levels (Dambach et al. 2005, Peters 2005). These *in vitro* preclinical tests are sensitive and functional (i.e., they generate conclusive results), but they ignore the importance of synergy between physiological systems.

Toxicogenomics is becoming an increasingly popular tool to identify the genetic basis of toxicity and to assist with safer drug development. Microarray analysis of drug exposure is mainly conducted using *in vitro* models but has been applied to some animal models. Microarray analyses conducted in an *in vitro* model has been criticized for ignoring the physiological interactions within an organism (Yang et al. 2004). These analyses are broad and sensitive, sometimes revealing thousands of differentially

expressed genes between treated and control tissues. However, the data generated from such analyses are nonfunctional because they generate inconclusive data; this is because it remains unclear if these expression changes are beneficial, neutral, or deleterious. In order to determine if the differential gene expression caused by the NCE is deleterious, these data are compared to microarray data of known toxicants (Yang et al. 2004), but because the reference database is largely incomplete, the problem remains in determining the functional consequences of differential gene expression caused by NCEs.

The objectives of *in vivo* preclinical trials are to determine if a NCE induces carcinogenicity, mutagenicity, reproductive toxicity, and teratogenicity (Kotsonis et al. 1985, Lipsky and Sharp 2001). NCEs are typically required to be tested in two different species, one rodent model, usually rat or mouse, and one nonrodent model, usually dog (Peters 2005). Concordance of human toxicity predicted by rodent models is 43% while concordance in nonrodents are typically higher, 63%, but by testing in both classes of animal models, toxicity in humans can be predicted with 70% accuracy (Olson et al. 2000). Once a "no observable adverse effect level" (NOAEL) has been established, then a safe dose can be established for humans and then the NCE enters clinical trials.

1.3 Organismal Performance Assays (OPAs)

Failure to detect ADRs in preclinical studies may largely be due to the lack of assays that have all three criteria of being broad, sensitive, and functional. We have developed an assay, known as the Organismal Performance Assay (OPA), that is capable of revealing mammalian health consequences. OPAs utilize genetically diverse wild-derived mice (*Mus musculus*) where individuals from different treatments compete

directly with each other for limited resources in seminatural enclosures. Individual performance is measured in terms of Darwinian fitness (i.e., reproductive success) and components that lead to fitness (e.g., competitive ability and survival). OPAs are sensitive and broad because they challenge most physiological systems synergistically and in order for individuals to be successful (functional), high performance from these physiological systems is required (broad and sensitive). In several instances, OPAs have proven capable of detecting mammalian health declines with higher sensitivity than current *in vivo* and *in vitro* approaches. Pharmaceutical companies utilizing OPAs have the potential to dramatically increase the detection rate of harmful substances before these pharmaceutical move onto human testing.

All OPAs consist of seminatural enclosures that are ~ 30 m^2 and are divided into six territories by hardware mesh. These fences are easily scaled but add an element of spatial complexity. Each territory has free access to food and water (gallon poultry waterers) and contains multiple nesting sites. Within an enclosure, four of the territories are considered optimal. The optimal dark nest boxes consist of a large plastic storage bin (75 L) with one defendable 5 cm diameter entrance. Within the tubs, there are three standard shoebox mouse cages with nesting material and also contain a 5 cm diameter entrance. Optimal territories also have direct access to food, where individuals do not need to leave the nest box to feed. The suboptimal nest boxes consist of two light-exposed plantar boxes with wire tops. Each plantar box has two 5 cm diameter entrance holes, making them difficult to defend. Individuals inhabiting the suboptimal territories have to leave the nesting sites to feed at the adjacent feeding station. Collectively, the fences and different nesting sites created environmental complexity that wild mice used

to establish territories and social hierarchy. Unlike animals housed in cages (like those used in preclinical studies), mice competing in these stressful seminatural enclosures require high performance from most of their physiological systems to establish social dominance and become successful individuals.

1.4 Validation of OPAs

OPAs have previously been used to quantify fitness costs associated with several different types of treatments, including genetic inbreeding. Inbreeding depression causes health consequences by increasing the likelihood of expressing deleterious recessive alleles that are otherwise masked in heterozygous outbred animals (Meagher et al. 2000). Previous lab based experiments revealed a 10% reduction in litter size when parents were full siblings (Lynch 1977, Connor and Belucci 1979). However, these experiments were conducted on animals within a caged environment and therefore lacked natural stressors such as competition for mates and resources. OPAs were used to assess fitness degradations in inbred offspring resulting from full-sibling crosses. Inbred progeny were released into seminatural enclosures where they competed directly with outbred individuals. OPAs quantified health degradations of full-sibling inbred progeny and male offspring were found to have a 500% fitness decline when they were in direct competition with outbred males (Meagher et al. 2000). OPAs were then used to examine the fitness costs associated with a single generation of cousin-level inbreeding and was found to cause a 34% fitness in reduction (Ilmonen et al. 2008), an adverse effect that remained undetected in previous studies.

In addition to examining the negative effects associated with genetic inbreeding, OPAs have also been used to identify and quantify the deleterious effects of a selfish gene. The *t* complex has become a classic example of a selfish gene ever since its discovery nearly half a century ago. The *t* complex sabotages wild-type gametes and is inherited at a frequency of ~90%; however, in wild mouse populations, the *t* complex is detected at levels far below the expected rate (Lewontin 1968). Consequently, it was predicted that some form of selection has prevented the *t* haplotype from going to fixation within wild mouse populations. OPAs were used to evaluate this phenomenon where wild type individuals (+/+ homozygotes) were in direct competition with heterozygotes (+/*t*) and *t*/*t* homozygotes (*t*/*t* causes male sterility or embryonic mortality). OPAs revealed a strong selection against *t* bearing heterozygotes in both males and females. Heterozygous males were less dominant than wild type males and because females prefer to mate with dominant males at a higher frequency, this selection lead to the *t* complex being inherited at a rate ~49% less than what was expected in the wild (Carroll et al. 2004).

Most recently, OPAs have been used to quantify the impacts of added sugar at human relevant levels. Added sugar consumption has been correlated with a number of diseases such as obesity, type-2 diabetes, metabolic syndrome, hypertension, and fatty liver disease (Bray et al. 2004, Gross et al. 2004, Dhingra et al. 2007, Ouyang et al. 2008, Fung et al. 2009) In the early 1970s, high fructose corn syrup (HFCS) became progressively incorporated into processed foods such as soft drinks and desserts and by 1985, HFCS accounted for approximately 35-40% of sweeteners in food (Elliott et al. 2002, Bray et al. 2004). Two OPA experiments were used to assess fitness impacts of added sweeteners at human relevant levels. The first experiment addressed whether

fitness was impacted when exposed to 25% Kcals coming from HFCS when compared with starch-fed individuals. HFCS-fed females experienced a twofold increased in mortality while males occupied ~25% fewer territories and had ~25% fewer offspring (Ruff et al. 2013). The second experiment compared the health effects between two types of added sugar, HFCS and sucrose (table sugar). In this experiment, HFCS fed females experienced approximately a twofold increase in mortality and experienced reduced reproduction of ~25% when compared to sucrose-fed females (Ruff et al. *in review*). In all of these studies (genetic inbreeding, selfish gene, and added sugar), OPAs found substantial deleterious effects that were missed by current methodologies.

The following chapters represent the first application of OPAs to detect and quantify fitness impacts associated with pharmaceutical drug exposure. We chose three pharmaceuticals that have already been recalled after market release (Vioxx® and Baycol®) or that remain on the market with warnings, while adverse effects are controversial (Paxil®). By testing pharmaceuticals that are already known to cause or suspected of causing adverse health consequences, we can validate OPAs as a sensitive technique and suggest they be implemented during preclinical trials.

1.5 Chapter Summaries

In Chapter 2, OPAs were used to evaluate the safety of cerivastatin (Baycol®, Bayer), a pharmaceutical belonging to the statin drug class, which aim to lower low-density lipoprotein (HDL) cholesterol. Cerivastatin was recalled three years after market release due to unacceptable adverse health consequences (Wooltorton 2001). During preclinical assessments, cerivastatin was deemed safe after it was found to have a similar

toxicological profile of other statins (Keutz and Schluter 1998). Despite the large investment of resources into preclinical and clinical trials, cerivastatin was found to cause rhabdomyolysis (breakdown of skeletal muscle causing kidney toxicity) at a rate of 16-80 times greater than other statins and was responsible for > 50 human fatalities (Furberg and Pitt 2001, Pasternak et al. 2002). Cerivastatin safety was assessed by OPAs and revealed numerous health consequences associated with previous exposure (from *in utero* into early adulthood) at a dose 27 times human therapeutic dose, but still within range of preclinical studies (Keutz and Schluter 1998). Within OPAs, cerivastatin-exposed males experienced threefold increase in mortality, sired 41% fewer offspring, and occupied 63% fewer territories when compared with controls. Cerivastatin-exposed females also experienced reduced reproduction where they had 25% fewer offspring than control females. These fitness declines escaped the detection of current preclinical methodologies and thus suggest that fitness assays like OPAs could be a powerful tool for safety testing during pharmaceutical development.

In Chapter 3, OPAs were used to assess the safety paroxetine (Paxil®, GlaxoSmithKline), a pharmaceutical currently available on the market. Paroxetine belongs to the selective serotonin reuptake inhibitors (SSRIs) drug class that are prescribed to treat cases of depression, anxiety, and other mood disorders. Paroxetine was approved for market use in the early 1990s and was later suspected of causing congenital malformations in babies born to mothers who took the drug during pregnancy; however, many studies are still controversial (Kulin et al. 1998, Williams and Wooltorton 2005, Bérard et al. 2007, Källén and Otterblad Olausson 2007, Diav-Citrin et al. 2008, Ellfolk and Malm 2010). In 2005, the FDA requested that paroxetine be labeled as a class D drug

(positive evidence of human fetal risk) and issued a warning that paroxetine exposure in the first trimester may potentially cause birth defects (FDA 2005). During typical preclinical trials, with doses 2.5-8-fold higher than human therapeutic dose, paroxetine caused infertility in rats but was not found to be teratogenic. In our experiment, exposure of paroxetine was administered near human therapeutic doses via food and occurred *in utero* and continued until early adulthood. Exposure ceased when animals were released into OPAs, thus any adverse health consequences detected would have been from previous exposure. OPAs revealed that paroxetine exposure decreased male reproductive success by 44%, decreased male competitive ability by 53%, and increased mortality by 2.5-fold. Paroxetine-exposed females had reduced reproduction of 65% at the initial OPA time point but rebound at later time points, suggesting that the negative effects of the drug wore off in females. These data indicate that OPAs provide superior sensitivity compared to conventional approaches and indicate OPAs could be a useful tool in safety testing during pharmaceutical development.

In Chapter 4, we evaluated a particularly interesting result we detected in the previously described experiment regarding paroxetine exposure. Litters from paroxetine breeders were significantly biased towards female offspring while litter size was unaffected when compared to control breeders. Sex allocation theory suggests that parents that facultatively adjust their sex ratio and/or invest more resources into one sex over the other can increase their lifetime reproductive success. Trivers and Willard (1973) suggest that in polygynous mating systems, high-quality mothers will produce high-quality sons, who will outreproduce their sisters of the same condition, while the opposite is true. The cost of reproduction hypothesis suggests that differential resource

investment will increase the prospect that one sex will survive and reproduce (Cockburn et al. 2002); both of these ideas are centralized around increasing individual lifetime fitness by favoring the sex of the offspring that has the highest prospect of surviving and passing on genetic material into future generations. In this experiment, we examined whether paroxetine-exposed mothers saved energy (in terms of offspring wean weight) by producing fewer sons. We then experimentally tested whether female offspring from female biased litters out reproduced the males from the same litter, when competing directly against individuals from unbiased (control) litters under seminatural conditions. Male offspring born to control mothers (unbiased litters) weighed significantly more than female offspring; however, no differences in wean weights were detected between male and female offspring from paroxetine-exposed mothers (biased litters). This suggests that paroxetine-exposed mothers did not save energy by skewing their litters towards female offspring because males did not cost more to raise than females. When offspring were in seminatural enclosures, we found that female offspring from paroxetine-exposed mothers had significantly more offspring compared to their brothers relative to controls. These data suggest that paroxetine-exposed mothers were successful at increasing their lifetime reproductive success by skewing litters towards females and is the first experimental evidence of determining the reproductive success of progeny from biased litters that provides support for the Trivers-Willard hypothesis.

In Chapter 5, OPAs were used to assess the safety of rofecoxib (Vioxx®, Merck), a selective nonsteroidal anti-inflammatory drug (NSAID). Rofecoxib was on the market from 1999 until 2004 and was recalled after being linked with causing heart attacks in 27,000 patients who took the drug. The deleterious effects associated with rofecoxib

escaped the detection of OPAs. No differences were detected in male fitness, where control and rofecoxib-exposed males had equal reproductive success, survival, and competitive ability. Rofecoxib exposed females, however, had 40% more offspring than control females. This is only the second study to which OPAs failed to detect adversity, and therefore, OPAs would still be useful if implemented drug development.

1.6 References

Bérard, A., E. Ramos, E. Rey, L. Blais, M. St-André, and D. Oraichi. 2007. First trimester exposure to paroxetine and risk of cardiac malformations in infants: the importance of dosage. Birth Defects Research. Part B, Developmental and reproductive toxicolocy 80:18–27.

Bray, G. A., S. J. Nielsen, and B. M. Popkin. 2004. Commentary Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. The American Journal of Clinical Nutrion 79:537–43.

Brimecombe, J. C., G. E. Kirsch, and A. M. Brown. 2009. Test article concentrations in the hERG assay: losses through the perfusion, solubility and stability. Journal of pharmacological and toxicological methods 59:29–34.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Cockburn, A., S. Legge, and M. Double. 2002. Sex ratios in birds and mammals: can the hypotheses be disentangled? *in* Sex Ratios Concepts and Methods. Pages 266–286 (I. Hardy, Ed.). Cambridge University Press, Cambridge, UK.

Connolly, H. M., J. L. Crary, M. D. McGoon, D. D. Hensrud, B. S. Edwards, W. D. Edwards, and H. V Schaff. 1997. Valvular heart disease associated with fenfluramine-phentermine. The New England journal of medicine 337:581–8.

Connor, J., and M. Belucci. 1979. Natural selection resisting inbreeding depression in captive wild house mice (Mus musculus). Evolution 33:929–40.

Dambach, D., B. Andrews, and F. Moulin. 2005. New technologies and screening strategies for hepatotoxicity: use of in vitro models. Toxicologic pathology 33:17–26.

Dhingra, R., L. Sullivan, P. Jacques, T. Wang, C. Fox, J. Meigs, R. D'Agostino, J. Gaziano, and R. Vasan. 2007. Soft drink consumption and risk of developing cardiometabolic risk factors and the metabolic syndrome in middle-aged adults in the community. Circulation 116:480–8.

Diav-Citrin, O., S. Shechtman, D. Weinbaum, R. Wajnberg, M. Avgil, E. Di Gianantonio, M. Clementi, C. Weber-Schoendorfer, C. Schaefer, and A. Ornoy. 2008. Paroxetine and fluoxetine in pregnancy: a prospective, multicentre, controlled, observational study. British journal of clinical pharmacology 66:695–705.

Ellfolk, M., and H. Malm. 2010. Risks associated with in utero and lactation exposure to selective serotonin reuptake inhibitors (SSRIs). Reproductive toxicology 30:249–60.

Elliott, S., N. Keim, J. Stern, K. Teff, and P. Havel. 2002. Special Article Fructose, weight gain, and the insulin resistance syndrome. The American journal of clinical nutrion 76:911–922.

Faich, G., and R. Moseley. 2001. Troglitazone (Rezulin) and hepatic injury. Pharmacoepidemiology and drug safety 10:537–47.

FDA. 2005. Public Health Advisory: Paroxetine. Accessed March 2014. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/PublicHealthAdvisories/ucm051731.htm.

Fung, T., V. Malik, K. Rexrode, J. Manson, W. Willett, and F. Hu. 2009. Sweetened beverage consumption and risk of coronary heart disease. The American journal of clinical nutrion 89:1037–42.

Furberg, C., and B. Pitt. 2001. Withdrawal of cerivastatin from the world market. Current controlled trials in cardiovascular medicine 2:205–7.

Gross, L., L. Li, E. Ford, and S. Liu. 2004. Increased consumption of refined carbohydrates and the epidemic of type 2 diabetes in the United States: an ecologic assessment. The American journal of clinical nutrion 79:774–9.

Ilmonen, P., D. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Källén, B., and P. Otterblad Olausson. 2007. Maternal use of selective serotonin re-uptake inhibitors in early pregnancy and infant congenital malformations. Birth defects research. Part A, Clinical and molecular teratology 79:301–8.

Keutz, E. Von, and G. Schluter. 1998. Preclinical safety evaluation of cerivastatin, a novel HMG-CoA reductase inhibitor. The American journal of cardiology 82:11J–17J.

Kotsonis, F., D. Dodd, B. Regnier, and F. Kohn. 1985. Preclinical toxicology profile of misoprostol. Digestive diseases and sciences 30:142S–146S.

Kulin, N., A. Pastuszak, S. Sage, B. Schick-Boschetto, G. Spivey, M. Feldkamp, K. Ormond, D. Matsui, K. Stein-Schechman, L. Cook, J. Brochu, M. Rieder, and G. Koren. 1998. Pregnancy outcome following maternal use of the new selective serotonin reuptake inhibitors: a prospective controlled multicenter study. The journal of the American medical association 279:609–10.

Lewontin, R. 1968. The effect of differential viability on the population dynamics of t alleles in the house mouse. Evolution 22:262–73.

Lipsky, M., and L. Sharp. 2001. From idea to market: the drug approval process. The journal of the American board of family practice 14:362–7.

Lynch, C. 1977. Inbreeding effects upon animals derived from a wild population of Mus musculus. Evolution 31:526–37.

Meagher, S., D. Penn, and W. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the national academy of sciences of the United States of America 97:3324–9.

Miller, H. 2012. How is the FDA really doing? Genetic Engineering and Biotechnology News 32:6–8.

Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes, W. Bracken, M. Dorato, K. Van Deun, P. Smith, B. Berger, and A Heller. 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regulatory toxicology and pharmacology 32:56–67.

Ouyang, X., P. Cirillo, Y. Sautin, S. McCall, J. Bruchette, A. Diehl, R. Johnson, and M. Abdelmalek. 2008. Fructose consumption as a risk factor for non-alcoholic fatty liver disease. Journal of hepatology 48:993–9.

Pasternak, R., S. Smith, C. Bairey-merz, S. Grundy, J. Cleeman, and C. Lenfant. 2002. ACC/AHA/NHLBI Clinical Advisory on the Use and Safety of Statins. Journal of the American college of cardiology 40:567–72.

Peters, T. S. 2005. Do preclinical testing strategies help predict human hepatotoxic potentials? Toxicologic pathology 33:146–54.

Ruff, J., S. Hugentobler, A. Suchy, M. Sosa, R. Tanner, L. Morrison, S. Gieng, M. Shigenaga, and W. Potts. (in review). Sugars within high fructose corn syrup reduce survival and fitness of female mice compared to surcrose. The Journal of the federation of American societies for experimental biology.

Ruff, J., A. Suchy, S. Hugentobler, M. Sosa, B. Schwartz, L. Morrison, S. Gieng, M. Shigenaga, and W. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

Schuster, D., C. Laggner, and T. Langer. 2005. Why drugs fail - A study on side effects in new chemical entities. Current pharmaceutical design 11:3545–59.

Trivers, R., and D. Willard. 1973. Natural selection of parental ability to vary the sex ratio of offspring. Science 179:90–2.

Williams, M., and E. Wooltorton. 2005. Paroxetine (Paxil) and congenital malformations. Canadian medical association journal 173:1320–1.

Wooltorton, E. 2001. Bayer pulls cerivastatin (Baycol) from market. Canadian medical association journal 165:632.

Yang, Y., E. Blomme, and J. Waring. 2004. Toxicogenomics in drug discovery: from preclinical studies to clinical trials. Chemico-biological interactions 150:71–85.

CHAPTER 2

CERIVASTATIN EXPOSURE DECREASES FITNESS, SURVIVAL, AND COMPETITIVE ABILITY OF HOUSE MICE LIVING UNDER SEMINATURAL CONDITIONS

2.1 Abstract

We utilized a novel toxicity assessment method known as the organismal performance assay (OPA) to evaluate the safety of cerivastatin, a drug withdrawn from the market due to numerous unforeseen toxicities and human fatalities. OPAs evaluate individual performance by utilizing genetically diverse wild mice (*Mus musculus*) that are assigned to treatment and control groups that compete amongst each other for limited resources in seminatural enclosures. Performance measures included reproductive success, survivorship, and male territorial acquisition. Within OPAs, cerivastatin-exposed females had 25% fewer offspring when compared with controls. Cerivastatin-exposed males weighed 10% less, occupied 63% fewer territories, sired 41% fewer offspring, and experienced a threefold increase in mortality when compared to controls. OPAs detected several fitness declines induced by cerivastatin, which were not detected with current preclinical toxicology assessments, indicating that fitness assays like OPAs could be a powerful tool for safety testing during pharmaceutical development.

2.2 Introduction

Statins belong to a well-known drug class that are prescribed to patients to reduce low-density lipoprotein (LDL) cholesterol (Sirvent *et al*., 2008). In 1998, cerivastatin (Baycol®, Lipobay®; Bayer, Leverkusen, Germany) was approved for market use, but was recalled in 2001 due to unacceptable health consequences (Wooltorton, 2001). Specifically, cerivastatin was linked to 385 nonfatal and 52 fatal cases of rhabdomyolysis (the breakdown of skeletal muscle leading to kidney toxicity) in the 700,000 patients taking the drug (Furberg and Pitt, 2001). Forty percent of fatalities were linked to a drug-drug interaction with gemfibrozil (Lopid; Pfizer). However, when taken alone, cerivastatin was still 16 to 80 times more likely to cause rhabdomyolysis than other statins when prescribed at the highest therapeutic dose of 0.8 mg per day (Furberg and Pitt, 2001; Pasternak *et al*., 2002).

Preclinical studies were conducted in several animal models to assess the effects of cerivastatin on teratogenicity, mutagenicity, carcinogenicity, and fertility (Keutz and Schluter, 1998). Adverse effects on the liver and muscle tissue were detected in rodent models, but cerivastatin was not teratogenic, mutagenic, nor did it cause infertility (Keutz and Schluter, 1998). Preclinical trials suggested cerivastatin to have a similar toxicological profile to that of other statins whose adverse effects were primarily in muscle tissue and the liver (Keutz and Schluter, 1998). These studies deemed cerivastatin safe, which supported moving cerivastatin into human clinical trials.

The case of cerivastatin is not unique because over the course of drug development, 73% of pharmaceuticals that pass the preclinical phase of testing fail during clinical trials (Lipsky and Sharp, 2001) and furthermore, 10% of all Food and Drug

Administration (FDA) approved pharmaceuticals fail after market release due to unforeseen adversity (Schuster *et al*., 2005). Not only is drug development fraught with risk, but it requires between 12-15 years of research and also typically costs companies $1.4 billion per drug (Miller, 2012). In addition to human cost in pain, suffering, and loss of life, drugs that fail after market release can cause pharmaceutical companies substantial losses due to litigation fees; for example, Bayer's settlement for cerivastatin was approximately $1.2 billion (Campbell *et al*., 2010). One potential cause of numerous drugs failing after years of research is the inability of current preclinical methodologies to detect adverse drug reactions (ADRs) without ambiguous interpretation; many toxicity assessment methods detect differences in parameters, but provide little information on whether observed changes are biologically adverse, beneficial, or neutral.

We have developed a novel toxicity assay, known as the organismal performance assay (OPA), which provides an unrivaled combination of breadth, sensitivity, and unambiguous adversity information, which will make it valuable during preclinical testing. OPAs utilize genetically diverse wild-derived mice (*Mus musculus*) that compete amongst each other for limited resources, such as food and mates, in a seminatural environment. Treatment and control individuals compete directly and the performance of individuals is measured in terms of Darwinian fitness (i.e., reproductive success) and the components that influence fitness (i.e., survival and male competitive ability). The sensitivity of the OPA derives from the fact that wild mice under social competition allows small changes in behavior or physiological performance and otherwise cryptic effects of toxicity to be manifested as measureable negative outcomes, such as relegation to inferior habitat and reduced reproduction and survival. Consequently, any degradation

in health caused by a treatment will be detectable by the inability of mice to perform comparable to controls with whom they compete. OPAs have been used to quantify the adverse effects of a broad array of treatments, including both genetic and nutritional manipulations (Meagher *et al*., 2000, Carroll *et al*., 2004, Ilmonen *et al*., 2008, Ruff *et al*., 2013); in all of these studies, OPAs found substantial deleterious effects that were missed by available *in vitro* and *in vivo* approaches.

Here we use OPAs to determine if cerivastatin exposure causes health and fitness declines in mice. Our findings will address the current lack of information concerning consequences of cerivastatin exposure at an organismal level. If cerivastatin exposure adversely affects any physiological system, we predict that exposed individuals will suffer survival and reproduction declines and males will be less dominant relative to controls. If OPAs are successful in detecting cerivastatin-induced adversity, these results will provide evidence that OPAs would be powerful tool if implemented during preclinical studies, where previous preclinical methodologies deemed cerivastatin to be safe.

2.3 Materials and Methods

2.3.1 Animals. Genetically diverse, wild-derived house mice were used in this experiment as they have retained behavioral traits that allow them to function successfully in seminatural environments unlike mice from inbred lines (Nelson *et al*., 2013). Individuals used in this experiment were from the 12th generation of the colony initially described in Meagher *et al*., (2000). Consanguinity was assessed during the 11th generation and found to be comparable to wild populations (Cunningham *et al*., 2013).

Animals were kept on a 12:12 h light:dark cycle and provided with food and water *ad libitum*. All procedures and protocols were approved by IACUC at the University of Utah.

2.3.2 Cerivastatin exposure. Dosing was achieved by incorporating 1.5 g of cerivastatin (Sequoia Research Products, Pangbourne, United Kingdom; Molecular formula: $C_{26}H_{33}FNNaO_5$) into 50 kg of standard rodent chow (TD.130006; Harlan Teklad, Madison, WI). As wild mice eat ~3 g per day, individuals received ~0.09 mg of cerivastatin daily, or ~4.5 mg/kg/day, assuming the average mouse weighs 20 g (Reagan-Shaw *et al*., 2008). Using a metabolic rate conversion factor, this is equivalent to a human dose of ~0.36 mg/kg/day, or a daily dose of ~21.6 mg, assuming the average human weighs 60 kg (Reagan-Shaw *et al*., 2008). Cerivastatin was prescribed up to a daily dose of 0.8 mg (Furberg and Pitt, 2001). Our mouse dose was approximately 27 times higher than human therapeutic doses, but still within the range of preclinical testing (Keutz and Schluter, 1998).

Forty breeding pairs were randomly and equally assigned to either cerivastatin or control treatments. Females and males were kept separate for one week prior to breeding. Cerivastatin females were exposed seven days prior to pairing and males were exposed six days prior. Breeders were kept together until they produced a maximum of three litters. At 28 days, pups were weaned and separated into same sex sibling cages. Data collected during weaning consisted of litter size, pup weights, and sex. Exposure continued until pups became adults and were released into enclosures. This duration of cerivastatin exposure maximized the ability of OPAs to detect health consequences because once animals were released into the seminatural enclosures, they were all fed the

control diet as currently, we are unable to keep animals on their respective diets while they are free ranging during OPAs. While control animals could have been switched to the cerivastatin treatment, switching the cerivastatin-exposed animals to the control diet was the most conservative approach of detecting fitness declines, as any differential fitness represented as accumulated damage due to the diet, not the current health degradation.

2.3.3 OPA enclosures. Enclosures have been extensively described in Ruff et al., (2013). Briefly, enclosures were ~30 m^2 and contained four optimal and two suboptimal territories (Figure S1). Optimal territories contained dark nesting sites, whereas suboptimal nesting sites were exposed to light. Boundaries between territories were established with wire mesh fencing that was easily climbed, but provided spatial complexity. Each territory contained access to food and water and a 12:12 h light:dark cycle was maintained.

Five independent OPA populations were established, each consisting of eight males and 12-16 females for a total of 116 (40 males, 76 females) individuals that are referred to as population founders. The space and animal density is consistent with the range of observations from the wild (Sage, 1981). Half of the individuals of each sex were exposed to cerivastatin while the remaining half were from the control treatment. Males were on average 15.9 ± 3.9 (M±SD) weeks old and females were 16.7 ± 4.3 weeks old at the time of enclosure release. To prevent incidental breeding before territories were established, males were released into the enclosures with nonexperimental females. After one week, when territories were established, nonexperimental females were removed and replaced with experimental females. Male founders were unrelated at the cousin level or

above and each population consisted of one to three pairs of sisters, but the relatedness between females was balanced between treatments. Populations were maintained for 28 weeks.

2.3.4 Male competitive ability. One week prior to release, all animals received a passive integrated transponder (PIT) tag (TX1400ST, BioMark, Boise, ID) for identification purposes. PIT tag antennas were placed on top of feeding stations within each of the six territories in an enclosure and data from PIT tag readers (FS2001F-ISO, BioMark, Boise, ID) were continuously downloaded to a computer containing data-logging software (Minimon, Culver City, CA). Two sets of PIT tag readers and antenna were used to collect data on OPA populations and were rotated twice weekly among contemporaneous populations. A male was assigned as a territorial occupant when he triggered > 80% of all male PIT tag readings in a particular territory over a multiday reading frame. PIT tag data were collected on females, but were not analyzed due to the lack of knowledge concerning female dominance behavior.

2.3.5 Reproductive success. To identify reproductive success, offspring were removed every five weeks, sacrificed, and had a tissue sample collected for genetic analysis. These "pup sweeps" prevented the young from reaching sexual maturity, breeding with the founders, and confounding the reproductive data. The first pup sweep occurred at week eight; the oldest offspring during this sweep were five weeks old as the gestation period is three weeks. A total of 1,668 samples were collected with an average of 333.6 ± 105.3 (M ± SD) offspring per population.

In four of the five populations, reproductive success was determined on a population-level by examining sex-specific allelic variants described in Meagher *et al*.,

(2000). Briefly, in each population, founding individuals of each treatment were selected based on nonoverlapping allelic variants on the Y-chromosome for males and on the mitochondrial genome for females. To control for confounding effects, such as segregating genes linked with the markers, marker assignment was balanced across treatments among populations. Mitochondrial genotypes were assessed in 1,280 samples (four of five populations) and obtained for 100% of offspring. Y-chromosome genotypes were assessed in all populations to determine reproductive output of males. Of the 1,668 offspring, 818 Y- chromosome genotypes were obtained, suggesting that 98% of all males were typed if the sex ratio was 1:1; successfully genotyping nearly all of the progeny allowed for a thorough analysis of founder reproductive success.

Female reproductive success was determined by parentage analysis using multiple microsatellite loci in the one population not typed with the above method to gain more knowledge on individual founder reproductive success for another study. These data were converted to population level readouts (i.e., number of pups per treatment rather than number of pups per individual) and combined with the mitochondrial data for analysis. Between 5 and 11 autosomal microsatellite loci were amplified, scored, and analyzed in a stepwise fashion. Loci used were: d1mit449, d3mit22, d3mit312, d3mit333, d5mit139 d6mit138, d9mit251, d12mit277, d14mit128, d17mit62, and d19mit110. Primer sequences were obtained from the Mouse Genome Informatics website, The Jackson Laboratory, Bar Harbor Maine (http://www.informatics.jax.org/ accessed March 2014). Primers were fluorescently tagged with either CY-5 or CY-3 dye. PCR products were run on a 14” x 17”, 6.25% denaturing acrylamide gel at 40 W for three to seven hours (locus

dependent). Gels containing tagged DNA fragments were imaged on a Typhoon Scanner 8600 with ImageQuant software (Amersham Biosciences, Piscataway, NJ).

Parentage was assigned by using Cervus 3.0 (Kalinowski *et al*., 2007). Allele frequencies were calculated using the genotypes of all candidate mothers and fathers and all offspring within the population. Simulations were run at 10,000 cycles with an error rate of 1%. Assigned parents were accepted when the trio confidence of mother, father, and offspring was 95%. With this rule, 97% offspring (375 of 388) were assigned to parent pairs.

2.3.6 Survivorship. Survivorship was assessed by daily noninvasive health checks and by extensive checks during pup sweeps. These extensive checks were conducted every five weeks as to not disrupt territoriality formation that increases infanticidal behavior. Research personnel only entered the pens to rotate PIT-tag readers, fill feeders, freshen waterers, and conduct pup sweeps. Dead founders were removed and identified by PIT tags. The date of death was estimated based upon the condition of the corpse. Animals that were dead long before discovery were given a death date half way in between the last PIT tag read and the date the individual was found.

2.3.7 Statistical analyses. Linear mixed models (LMMs) were used to analyze wean weight and body weight in OPAs as these data are continuous and normally distributed; therefore, t-values were calculated to inform p-values in conjunction with degrees of freedom. Generalized linear mixed models (GLMMs) were used to analyze litter size, reproductive success, and male competitive ability as these data are discrete counts and conform to either a Poisson distribution (litter size and reproductive success) or binomial distribution (male competitive ability); therefore, z-scores (the number of

standard deviations apart two population means are) were calculated to inform p-values. To use a GLMM with a Poisson distribution, data are logarithmically transformed, thus standard errors (SE) are asymmetric as values reported in the results section have been back-transformed. Both LMMs and GLMMs were conducted in R 3.0.2 using the lme4 library (Bates *et al*., 2014, R Development Core Team, 2013). P-values were calculated for LMMs with the Swatterthwaite approximation under the lmerTest function (Kuznetsova *et al.*, 2013). Cox proportional hazard models (PH) were used for survivorship; therefore, χ^2 were calculated to inform p-values in conjunction with degrees of freedom (JMP 9.0.3, SAS institute Inc., Cary NC). A complete description of statistical analyses can be found in the supplementary information.

2.4 Results

2.4.1 Breeding cage measurements. No difference was detected in wean weight of cerivastatin-exposed first litters (model intercept) when compared with controls (LMM; $n = 27$, female offspring, $t = 0.68$, $p = 0.51$; $n = 29$, male offspring, $t = 0.71$, $p = 0.48$). Cerivastatin-exposed females weighed an average of 11.18 ± 0.55 g (M ± SE) and controls weighed 10.81 ± 0.35 g. Cerivastatin-exposed males weighed an average of 12.36 ± 0.53 g and controls weighed an average of 11.98 ± 0.34 g. No effect of time (LMM; female, $t = 1.30$, $p = 0.22$; male, $t = 1.67$, $p = 0.11$) or time by treatment occurred (LMM; female, $t = -0.81$, $p = 0.43$; male, $t = -0.56$, $p = 0.58$). For a complete readout of mixed model results for breeding cage data (litter size and wean weight), see Table S2.1.

Litter size was not affected by treatment in first litters (model intercept of GLMM; $n = 30$, $z = -0.04$, $p = 0.97$). Cerivastatin-exposed breeders produced 3.51 (SE

+0.83, −0.67) pups. Reported SEs are asymmetric as they were back-transformed from logarithmic data (see statistical analyses). Control breeders produced 3.49 (+0.50, −0.44) pups in their first litter. No effect of time (GLMM; $z = 0.10$, $p = 0.92$) or time by treatment occurred (GLMM; $z = 1.00$, $p = 0.32$).

2.4.2 OPA measurements. Cerivastatin exposure affected body weight of founders within OPAs. A marginally significant trend was detected in which cerivastatin-exposed females weighed 5% less than control females at week zero (model intercept of LMM; $n = 76$, $t = -1.74$, $p = 0.09$; Figure 2.1A). Cerivastatin-exposed females weighed an average of 18.69 ± 0.69 (M ± SE), whereas control females weighed on average 19.89 ± 0.78 g. Females from both treatments gained weight over time, presumably due to pregnancy (LMM; $t = 10.14$, $p < 0.0001$), but no interaction between time and treatment was detected (LMM; $t = 0.22$, $p = 0.83$), suggesting that the trend in reduced weight experienced by cerivastatin-exposed females at the intercept continued throughout the study. For a complete readout of mixed model results for OPA weight measures data, see Table S2.2.

Cerivastatin-exposed males weighed 10% less than controls at week zero (model intercept of LMM; $n = 40$, $t = -3.66$, $p < 0.0001$; Figure 2.1B). Cerivastatin-exposed males weighed on average 19.30 ± 0.56 g (M ± SE), while control males weighed on average 21.35 ± 0.56 g. Males gained weight over time (LMM; $t = 9.53$, $p < 0.0001$); however, no time by diet interaction was detected (LMM; $t = 1.05$, $p = 0.30$), suggesting that cerivastatin-exposed males weighed less than controls throughout the study.

Male competitive ability was negatively affected by cerivastatin exposure. Control males were 2.5 times more likely to occupy a territory than cerivastatin-exposed

males at week three (model intercept). Control males occupied 49% of the territories, while cerivastatin-exposed males only occupied 18% territories, leaving 32% territories undefended (GLMM; $n = 5$, $z = -4.55$, $p < 0.0001$; Figure 2.2). The percent of undefended territories is not unusual because 2/6 (or 33%) of the territories are suboptimal and are often difficult to defend. No effect of time (GLMM; $z = 0.99$, $p = 0.32$) or time by treatment were observed (GLMM; $z = -1.12$, $p = 0.26$), indicating that cerivastatin-exposed males acquired fewer territories over the duration of the study. For a complete readout of mixed model results for OPA competitive ability, see Table S2.3.

Cerivastatin females experienced a 25% reduction in reproductive success when compared with controls (GLMM; $n = 5$, $z = -3.70$, $p < 0.001$; Figure 2.3A). Cerivastatin-exposed females had an average of 21.82 (SE +2.09, −1.91) offspring per population, while controls had an average of 30.63 (+5.40, −4.58) per population at week eight (model intercept). Females from both treatments had more offspring over time (GLMM; $z = 3.39$, $p < 0.001$), but no time by treatment effect was detected (GLMM; $z = 0.89$, $p = 0.38$), suggesting the reproductive deficiency of cerivastatin-exposed females at the intercept was consistent across the study. For a complete readout of mixed model results for OPA reproduction, see Table S2.4.

Cerivastatin exposure also negatively affected male reproductive success where cerivastatin-exposed males had 41% fewer male offspring than control males (GLMM; $n = 5$, $z = -4.37$, $p < 0.0001$; Figure 2.3B). Cerivastatin-exposed males sired an average of 9.14 (SE +1.31, −1.15) offspring per population, while controls sired an average of 16.42 (+3.29, −2.76) offspring per population at week eight (model intercept). Males from both treatments had more offspring over time (GLMM; $z = 2.58$, $p < 0.01$), but no effect of

time by treatment was detected (GLMM; $z = 0.49$, $p = 0.62$), suggesting that the reproductive deficiency experienced by cerivastatin-exposed males was consistent throughout the study.

Cerivastatin-exposed individuals experienced higher mortality than controls. Cerivastatin-exposed females had increased mortality ($n = 76$); however, the data were not analyzed due to overall low mortality; only three cerivastatin females died while 100% of control females survived (Figure 2.4A). Cerivastatin-exposed males experienced a mortality rate 3.0 times that of controls (PH; $n = 40$, $\chi^2 = 4.79$, $p = 0.03$; Figure 2.4B). Mortality rate did not differ in replicate populations (PH; $\chi^2 = 1.65$, $p = 0.80$).

2.5 Discussion

Cerivastatin-exposed individuals suffered health and fitness declines when compared with controls. Cerivastatin-exposed males weighed 10% less, occupied 63% fewer territories, sired 41% less offspring, and had a threefold increase in mortality compared to controls. Cerivastatin-exposed females experienced 5% trend in reduced weight and had significantly fewer (25%) offspring when compared with controls. All of these negative effects had been missed by previous preclinical testing, suggesting that OPAs detected important organismal health declines and if OPAs had been in use during preclinical testing, they could have prompted additional studies that might have prevented this drug from moving to clinical trials and the marketplace.

Cerivastatin-exposed males experienced a reduction in weight in OPAs, which may be attributable to muscle toxicity and degeneration. Statins in general are known to cause myotoxicity, but typically occur at a low rate while only 1-7% of patients taking

statins experience this side effect (Sirvent *et al*., 2008). Additional research has revealed that exercise can exacerbate statin-induced myotoxicity (Sinzinger and O'Grady, 2004). Seachrist and others (2004) assessed the effect of exercise on muscle weight in cerivastatin-exposed rats. They found that both quadriceps and psoas muscles weighed significantly less in the exercised cohort (Seachrist *et al*., 2005). Unlike a caged environment, mice inhabiting OPAs engage in regular physical activity and males compete with each other to obtain and defend territories. It is possible that the increase in physical activity caused myotoxicity in cerivastatin-exposed male mice. Less severe weight differences were detected in female mice and may be due to less physical activity, while devoting more energy into reproductive efforts (Gittleman and Thompson, 1988).

Cerivastatin-exposed males were less competitive and dominated fewer territories than their control counterparts. This result might be due to decreased body weight as males engage in physical competitions to obtain and defend territories. Larger body size has been shown to be beneficial in territorial acquisition and defense in mice [e.g., (van Zegeren, 1980; Krackow, 1993)]. Alternatively, the cerivastatin-exposed males may have experienced muscle fatigue that hindered their endurance, leading to decreased competitive ability. A third possibility is that cerivastatin-exposed males might have experienced decreased testosterone levels. Testosterone plays a large role in male behavior, especially competitive behaviors in mice (Zielinski and Vandenbergh, 1993). The synthesis of testosterone and other steroid hormones relies upon available cholesterol (Bohm *et al*., 2004), which cerivastatin reduces. Some epidemiological evidence exists to suggest statin use does indeed lower testosterone levels (Hyyppä *et al*., 2003; Corona *et*

al., 2010); however, other studies have not found an association (Bohm *et al*., 2004; Hall *et al*., 2007) and thus, this issue is still controversial.

Cerivastatin-exposed males suffered reproductive declines when compared with controls. In a previous OPA study, dominant males were found to sire > 80% of offspring born in enclosures (Carroll *et al*., 2004); therefore, the differences detected in competitive ability are likely to explain, at least in part, the differences detected in reproduction. As cerivastatin-exposed males experienced a threefold increase in mortality, fewer males were reproducing and contributing to overall population-level reproduction. It is likely that other mechanisms contribute to reduced reproduction. For example, one study found that when male rats were exposed to atorvastatin (Lipitor®, Pfizer), at levels equivalent to 16 times human therapeutic dose, they had decreased sperm motility and spermatid head concentrations and increased sperm abnormalities (Pfizer, 2009), but cerivastatin induced reproductive impairments have not been detected in previous rodent studies (Keutz and Schluter, 1998).

Cerivastatin-exposed females had 25% fewer offspring in OPAs when compared to control females. The proximate mechanisms of reduced reproductive success are unknown; however, it is possible that reduced cholesterol levels caused by cerivastatin exposure may have contributed to these findings. Cholesterol is important for normal fetal development (Belknap and Dietschy, 1988) and is the precursor for several hormones, which play important roles in pregnancy and lactation (Henck *et al*., 1998; Bohm *et al*., 2004). Furthermore, statins inhibit the synthesis of isoprenoids, which are important for maintaining cell function and normal development (Surani *et al*., 1983). Another possibility is that cerivastatin might have unknown side effects on reproduction

that are not associated with reduced cholesterol levels. Even though the proximate mechanisms are not fully understood, OPAs detected important organismal health declines, which is ultimately the most critical information acquired during pharmaceutical safety assessment.

Although fitness declines were detected in both sexes, cerivastatin-exposed males suffered greater adversity than did females. This is not the first OPA study to detect differential health consequences between sexes. In three previous OPA studies, males have suffered greater negative fitness impacts than females; these studies include inbreeding at the cousin level (Ilmonen *et al*., 2008), inbreeding at the sibling level (Meagher *et al*., 2000), and when animals were fed a moderate sugar diet (Ruff *et al*. 2013; Figure S2.2). Male mice might have been affected by the cerivastatin treatment more than females due to the competitive nature of males in a polygynous mating system. Since OPA studies generate the same outputs, relative fitness can be compared between treatments. Results from this study indicate that cerivastatin-exposed male fitness (i.e., reproductive success) is approximately 25% less than the male progeny of parents that were first cousins (Figure S2.2).

OPAs have detected adversity in mice exposed to cerivastatin producing results without ambiguous interpretation. Unlike proximate-level based approaches of toxicity assessment, OPAs allowed for quantification of adverse impacts at the organismal level, requiring no assumptions of mechanistic cause. Not only were severe fitness declines detected in both males and females when exposed to cerivastatin, these results are conservative as all animals within OPAs were placed on the control diet. Thus, all fitness declines were caused by previous exposure and the cumulative damage the exposure

caused and although the proximate mechanisms of these phenotypes are unknown, OPAs could have prompted additional studies that could have led to discovery of the underlying mechanisms, which will help in determining if the similar adverse consequences are likely to occur in humans. If this had occurred prior to public release, it could have saved considerable human suffering and financial losses.

2.6 Conclusion

OPAs were able to quantify adversities due to cerivastatin exposure that were missed during both preclinical and clinical trials. Unfortunately, the failure of available screening methods to reveal toxicity led to market availability of cerivastatin for three years before it was found to cause unacceptable health consequences and thus was recalled. This study validates the ability of OPAs to reveal deleterious effects due to pharmaceutical exposures that were missed by current methodologies. OPAs achieve their sensitivity because they are unique in that hypotheses concerning specific mechanisms of action are not required; rather, whole organismal performance in its entirety is measured by OPAs within the context of a seminatural environment. Consequently, any toxicity that reduces performance of almost any physiological system (e.g., cardiorespiratory, metabolic, or neurological) is likely to be detected as the endpoint measures of this assay (such as reproductive success and survival) are integrated across physiological systems. Due to their sensitivity and lack of *a priori* assumptions, OPAs could prove useful in toxicity detection during pharmaceutical development, screening for potential environmental pollutants, and in quantifying the health consequences of nutritional substances.

2.7 Acknowledgments

We thank Dan Tripodi for suggesting that we apply OPAs to assess pharmaceutical safety; Michelle Bartlett, Stephanie Eddy, David Kircher, Elyse Schwab, Mirtha Sosa, Paige Wheatley, and Chloe Young for data collection and genotyping; and Ahmed Bwika and Jon Gale for animal care. This work was supported by the University of Utah's Technology Commercialization Program and was partially supported by the National Science Foundation [grant number DEB 09-18969 to W.K.P.]; by the National Institute of Health [grant number R01-GM109500 to W.K.P.]; and by the National Science Foundation GK-12 Educational Outreach Fellowship [grant number DGE 08-41233 to S.M.G.].

2.8 References

Bates, D., M. Maechler, B. Bolker, S Walker, R. Christensen, and H. Singmann. 2014. lme4: linear mixed-effects models using S4 classes. R package version:1.1-5.

Belknap, W. M., and J. M. Dietschy. 1988. Sterol synthesis and low density lipoprotein clearance in vivo in the pregnant rat, placenta, and fetus. Sources for tissue cholesterol during fetal development. The Journal of clinical investigation 82:2077–85.

Bohm, M., W. Herrmann, S. Wassmann, U. Laufs, and G. Nickenig. 2004. Does statin therapy influence steriod hormone synthesis. Clinical Research in Cardiology 93:43–8.

Campbell, J., A. Allen, and D. McIntyre. 2010. The ten worst drug recalls in the history of the FDA. Accessed April 2014. http://247wallst.com/investing/2010/12/10/the-ten-worst-drug-recalls-in-the-history-of-the-fda/.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Corona, G., V. Boddi, G. Balercia, G. Rastrelli, G. De Vita, A. Sforza, G. Forti, E. Mannucci, and M. Maggi. 2010. The effect of statin therapy on testosterone levels in

subjects consulting for erectile dysfunction. The journal of sexual medicine 7:1547–56.

Cunningham, C. B., J. S. Ruff, K. Chase, W. K. Potts, and D. R. Carrier. 2013. Competitive ability in male house mice (Mus musculus): genetic influences. Behavior genetics 43:151–60.

Furberg, C. D., and B. Pitt. 2001. Withdrawal of cerivastatin from the world market. Current controlled trials in cardiovascular medicine 2:205–7.

Gittleman, J. L., and D. Thompson. 1988. Energy Allocation in Mammalian Reproduction1. Integrative and Comparative Biology 28:863–75.

Hall, S. a, S. T. Page, T. G. Travison, R. B. Montgomery, C. L. Link, and J. B. McKinlay. 2007. Do statins affect androgen levels in men? Results from the Boston area community health survey. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 16:1587–94.

Henck, J. W., W. R. Craft, a Black, J. Colgin, and J. a Anderson. 1998. Pre- and postnatal toxicity of the HMG-CoA reductase inhibitor atorvastatin in rats. Toxicological sciences : an official journal of the Society of Toxicology 41:88–99.

Hyyppä, M. T., E. Kronholm, A. Virtanen, A. Leino, and A. Jula. 2003. Does simvastatin affect mood and steroid hormone levels in hypercholesterolemic men? A randomized double-blind trial. Psychoneuroendocrinology 28:181–194.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular ecology 16:1099–106.

Keutz, E. Von, and G. Schluter. 1998. Preclinical Safety Evaluation of Cerivastatin , a Novel HMG-CoA Reductase Inhibitor. The American Journal of Cardiology 82:11J–17J.

Krackow, S. 1993. The effect of weaning weight on offspring fitness in wild house mice (Mus musculus domesticus): A preliminary study. Ethology 95:76–82.

Kuznetsova, A., Brockhoff, P., and R. Christensen. 2013.lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4 package). R version 2.0-3.

Lipsky, M. S., and L. K. Sharp. 2001. From idea to market: the drug approval process. The Journal of the American Board of Family Practice 14:362–7.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

Miller, H. 2012. How is the FDA really doing? Genetic Engineering and Biotechnology News 32:6–8.

Nelson, A. C., J. W. Cauceglia, S. D. Merkley, N. A. Youngson, A. J. Oler, R. J. Nelson, B. R. Cairns, E. Whitelaw, and W. K. Potts. 2013. Reintroducing domesticated wild mice to sociality induces adaptive transgenerational effects on MUP expression. Proceedings of the National Academy of Sciences 110:19848–53.

Pasternak, R., S. Smith, C. N. Bairey-merz, S. M. Grundy, J. I. Cleeman, and C. Lenfant. 2002. ACC/AHA/NHLBI Clinical Advisory on the Use and Safety of Statins. Journal of the American College of Cardiology 40:567–572.

Pfizer. 2009. Lipitor (Atorvastatin Calcium) Tablets. http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/020702s057lbl.pdf.

R Development Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reagan-Shaw, S., M. Nihal, and N. Ahmad. 2008. Dose translation from animal to human studies revisited. The Federation of American Societies for Experimental Biology Journal 22:659–61.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

Sage, R. 1981. The Mouse in Biomedical Research. Pages 40–90 (H. Foster, J. Small, and J. Fox, Eds.). Academic Press, New York.

Schuster, D., C. Laggner, and T. Langer. 2005. Why Drugs Fail - A Study on Side Effects in New Chemical Entities. Current Pharmaceutical Design 11:3545–59.

Seachrist, J. L., C.-M. Loi, M. G. Evans, K. a Criswell, and C. E. Rothwell. 2005. Roles of exercise and pharmacokinetics in cerivastatin-induced skeletal muscle toxicity. Toxicological sciences : an official journal of the Society of Toxicology 88:551–61.

Sinzinger, H., and J. O. Grady. 2004. Professional athletes suffering from familial hypercholesterolaemia rarely tolerate statin treatment because of muscular problems 57:525–8.

Sirvent, P., J. Mercier, and A. Lacampagne. 2008. New insights into mechanisms of statin-associated myotoxicity. Current opinion in pharmacology 8:333–8.

Surani, M. a, S. J. Kimber, and J. C. Osborn. 1983. Mevalonate reverses the developmental arrest of preimplantation mouse embryos by Compactin, an inhibitor of HMG Co A reductase. Journal of embryology and experimental morphology 75:205–23.

The Jackson Laboratory. 2013. Accessed March 2014. http://www.informatics.jax.org.

Wooltorton, E. 2001. Bayer pulls cerivastatin (Baycol) from market. Canadian Medical Association Journal 165:632.

van Zegeren, K. 1980. Variation in aggressiveness and the regulation of numbers in house mouse population. Netherlands Journal of Zoology 30:635–770.

Zielinski, W. J., and J. G. Vandenbergh. 1993. Testosterone and competitive ability in male house mice, Mus musculus: laboratory and field studies. Animal Behaviour 45:873–91.

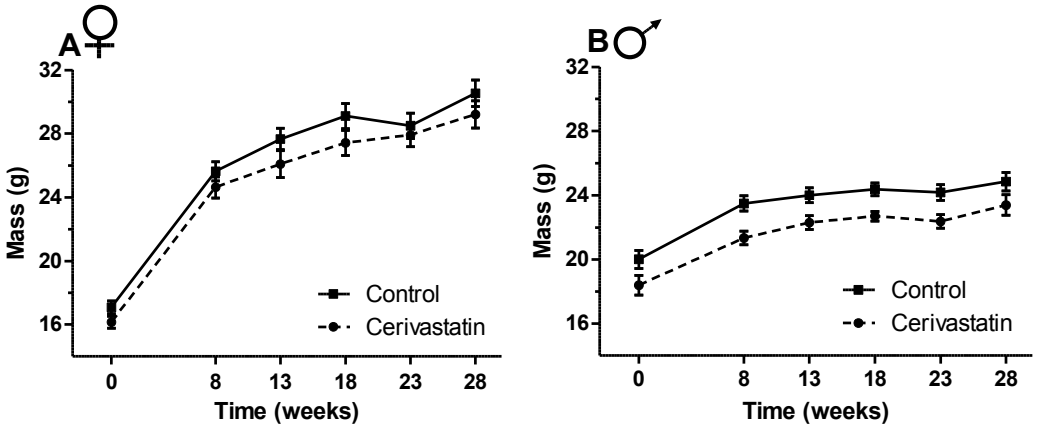

Figure 2.1. **Body weight of cerivastatin-exposed founders and controls over time in OPAs. A)** Cerivastatin-exposed females weighed 5% less than controls across the study, which was marginally significant [n =76 mice, *observations* = 483 (LMM; t = −1.74, p = 0.09). **B)** Cerivastatin-exposed males weighed significantly less (10%) than controls throughout the study [n = 40 mice, *observations* = 196 (LMM; t = −3.66, p < 0.001). Time point 0 is when animals were released into OPAs. Lines connect means and error bars represent standard error.

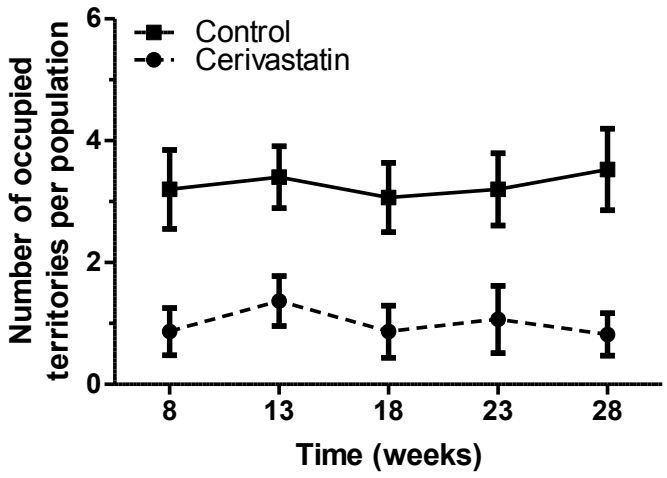

Figure 2.2. Competitive ability of cerivastatin-exposed males and controls.

Cerivastatin-exposed males occupied 63% fewer territories than control males throughout the duration of the experiment [n = 5 populations, *observations* = 130 (GLMM; $z = -4.55$, $p < 0.0001$)]. Within each enclosure, there were six available territories and a male was considered a territorial occupant if > 80% of all male PIT tag readings a particular territory belonged to him. On average, control males occupied 49% of the territories, cerivastatin-exposed males occupied 18% territories, and the remaining 32% territories were left undefended. The undefended territories still contain male mice; however, < 80% of male reads at this location belonged to a single individual. Points represent the mean number of territories of five populations. To aid in visualization, time points from five-week intervals have been pooled, except for the first time point consisting of eight weeks. For example, time point week eight consists of all data points collected from weeks 1-8; time points displayed at week 13 consists of all data points collected from weeks 9-13 and so on. Lines connect means of the five populations and error bars represent standard error.

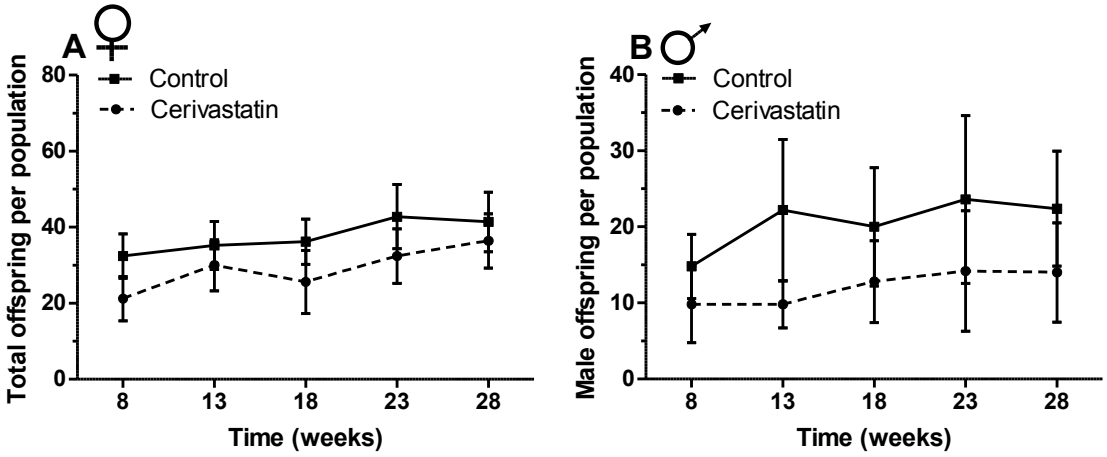

Figure 2.3. Reproductive success of cerivastatin-exposed animals and controls in OPAs. A) Cerivastatin-exposed females had 25% fewer offspring than controls throughout the study [n = 5 populations, *observations* = 50 (GLMM; $z = -3.70$, $p < 0.001$)]. **B)** Similarly, cerivastatin-exposed males sired 41% less offspring than controls [n = 5 populations, *observations* = 50 (GLMM; $z = -4.37$, $p < 0.0001$). Female reproduction is in terms of total offspring, as mitochondrial markers and multiple microsatellite markers were used. Male reproduction is in terms of male offspring, as Y-chromosomal markers were used. Lines connect means of the five populations at each time point for each sex and error bars represent standard error.

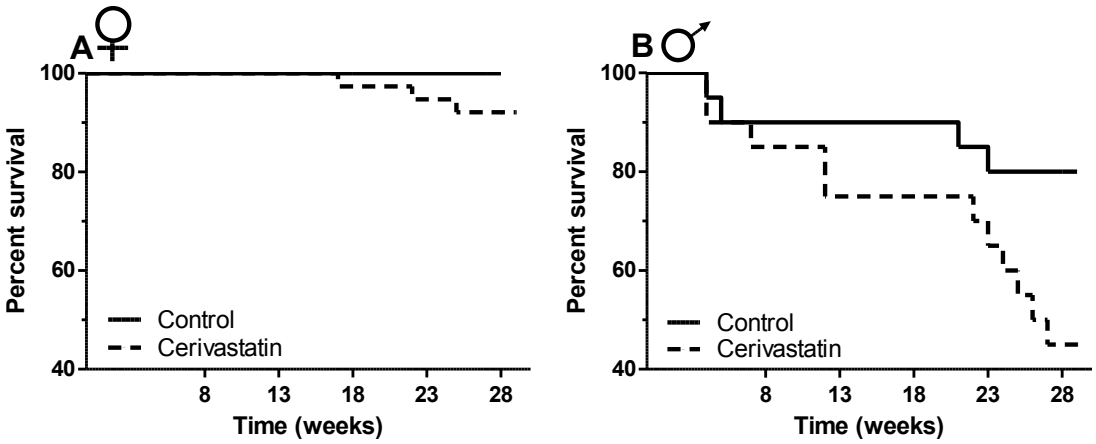

Figure 2.4. Survivorship of cerivastatin-exposed animals and control animals in OPAs. A) Cerivastatin-exposed females (*n* = 76) experienced reduced survival compared to controls; however, these data were not analyzed due to so few deaths: 100% survival of control females and three mortalities in the cerivastatin-exposed treatment. **B)** Cerivastatin-exposed males (*n* = 40) experienced a threefold increase in mortality than their control counterparts (PH; $\chi^2 = 4.79$, p = 0.03).

2.9 Supplementary Information

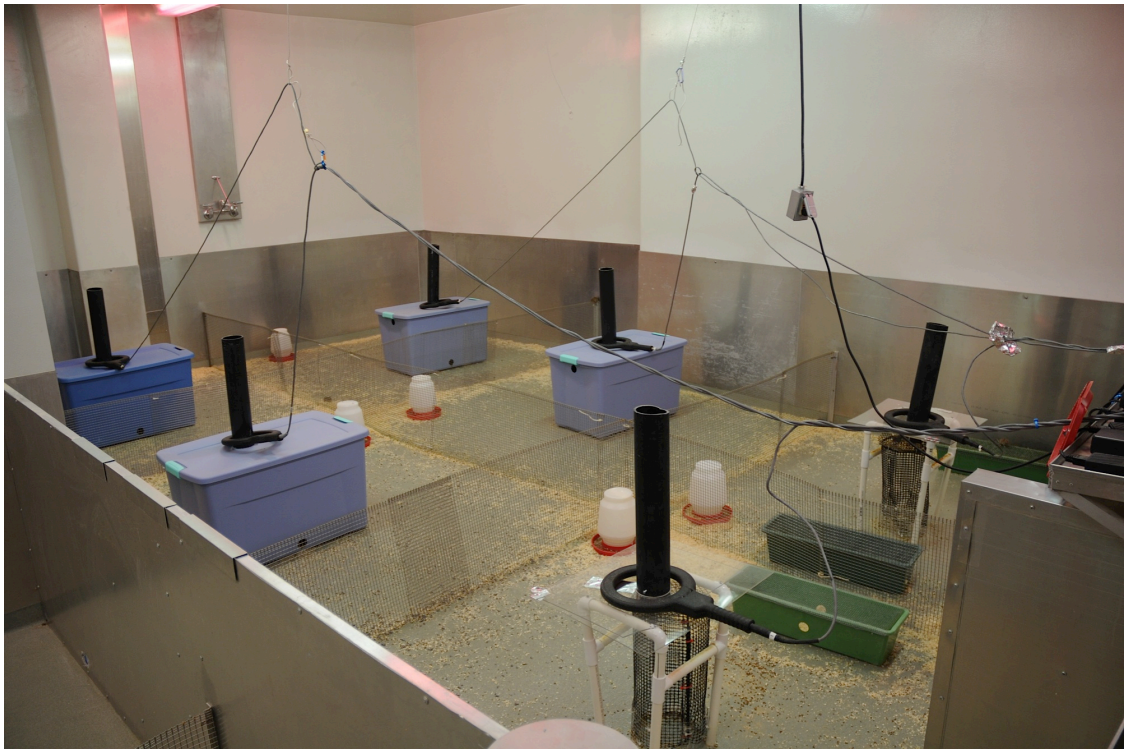

Figure S2.1. Image of a seminatural enclosure used in OPA experiments. Each enclosure is ~30 m^2 and contains 6 territories that are divided by wire mesh. The four optimal territories have the large blue bins, which contain multiple dark nesting sites and are defendable. The two suboptimal territories consist of light exposed nesting sites. Each territory contains food within the chimney-like structures and water (poultry waterers). PIT tag antennas (tennis racket like structures) are placed above each feeding site. Photograph courtesy Ben Sutter.

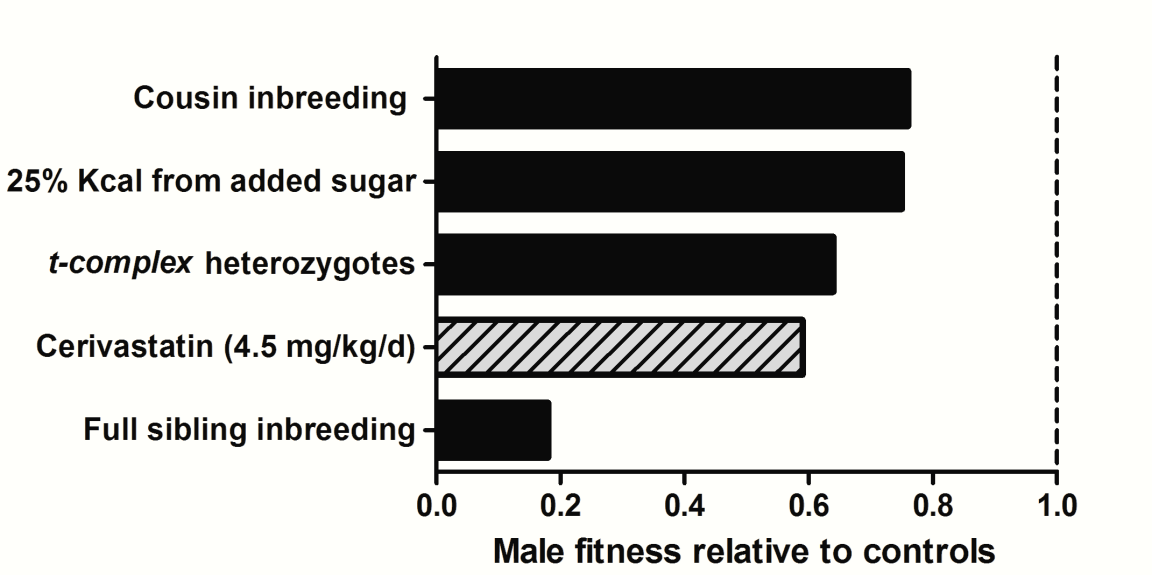

Figure S2.2. Comparisons of male fitness relative to control counterpart from published OPA experiments. Cerivastatin exposure had greater consequences on male fitness (reproductive success) when compared with cousin inbreeding, moderate sugar consumption, and a selfish gene but suffered less fitness consequences when compared to inbreeding at the sibling level (Meagher *et al*., 2000; Carroll *et al*., 2004; Ilmonen *et al*., 2008; Ruff *et al*., 2013). Figure adapted from Ruff *et al*., 2013.

Table S2.1. Mixed model results for litter size and wean weight.

Litter Size	GLMM with Poisson distribution and logarithmic link (groups = 30; observations = 68)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.0798	0.2824		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Litter 1)	1.2491	0.1335	9.3550	<0.0001***
Treatment (Cerivastatin)	0.0074	0.2124	0.0350	0.9720
Parity	0.0097	0.0988	0.0980	0.9220
Treatment (Cerivastatin)×Parity	0.1771	0.1763	1.0040	0.3150
Female Wean Weight	LMM (groups = 27; observations = 124)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	1.3423	1.1586		
Cage (Slope)	0.2616	0.5115		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Litter 1)	10.8131	0.3517	30.7420	<0.0001***
Treatment (Cerivastatin)	0.3676	0.5445	0.6750	0.5060
Parity	0.2769	0.2134	1.2980	0.2180
Treatment (Cerivastatin)×Parity	-0.2998	0.3725	-0.8050	0.4300
Male Wean Weight	LMM (groups = 29; observations = 136)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.9751	0.9875		
Cage (Slope)	0.3131	0.5596		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Litter 1)	11.9781	0.3381	35.4200	<0.0001***
Treatment (Cerivastatin)	0.3805	0.5378	0.7100	0.4840
Parity	0.4470	0.2675	1.6700	0.1110
Treatment (Cerivastatin)×Parity	-0.2711	0.4817	-0.5600	0.5790

*** Indicates a p value < 0.001.

Table S2.2. Linear mixed model results for body weight over time within OPAs.

Female Body Weight	LMM (group = 76; observations = 438)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	0.4176	0.6462		
Individual (Slope)	0.0209	0.1447		
Population (Intercept)	1.8593	1.3636		
Population (Slope)	0.0024	0.0489		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Week 0)	19.8851	0.7814	25.4490	<0.0001***
Treatment (Cerivastatin)	-1.1998	0.6917	-1.7350	0.0857 ·
Time	0.4309	0.0425	10.1400	<0.0001***
Treatment (Cerivastatin)×Time	0.0114	0.0519	0.2200	0.8263
Male Body Weight	LMM (group = 40; observations = 196)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	1.7187	1.3110		
Population (Intercept)	0.8147	0.9026		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Week 0)	21.3518	0.5636	37.8900	<0.0001***
Treatment (Cerivastatin)	-2.0497	0.5605	-3.6600	0.0005***
Time	0.1482	0.0156	9.5300	<0.0001***
Treatment (Cerivastatin)×Time	0.0246	0.0235	1.0500	0.2970

*** Indicates a p value < 0.001.

Table S2.3. Generalized linear mixed model results for male competitive ability over time.

Male Competitive Ability	GLMM with binomial distribution and logit link (group = 5; observations = 130)			
Random effects	*Variance*	*Std. Dev.*		
Population (Slope)	0.0570	0.2387		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 3)	-0.0260	0.2237	-0.1160	0.9080
Treatment (Cerivastatin)	-1.4625	0.3215	-4.5490	<0.0001***
Time	0.0125	0.0126	0.9890	0.3230
Treatment (Cerivastatin)×Time	-0.0235	0.0209	-1.1230	0.2620

· Indicates a p value <0.10, *** < 0.001.

Table S2.4. Generalized linear mixed model results for reproduction over time.

Female Reproduction	GLMM with Poisson distribution and logarithmic link (group = 5; observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.11328	0.3366		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 8)	3.4221	0.1622	21.1020	<0.0001***
Treatment (Cerivastatin)	-0.3394	0.0917	-3.7000	0.0002***
Time	0.0163	0.0048	3.3910	0.0007***
Treatment (Cerivastatin)×Time	0.0063	0.0071	0.8850	0.3764
Male Reproduction	GLMM with Poisson distribution and logarithmic link (group = 5; observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.1345	0.3668		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 8)	2.7980	0.1832	15.2690	<0.0001***
Treatment (Cerivastatin)	-0.5858	0.1341	-4.3660	<0.0001***
Time	0.0162	0.0063	2.5820	0.0098**
Treatment (Cerivastatin)×Time	0.0051	0.0103	0.4910	0.6233

** Indicates a p value < 0.01, *** < 0.001.

2.10 Statistical Analyses

Weight of offspring at weaning was analyzed with a linear mixed-effects model. This model assessed the effects of treatment, time, and the interaction of time and treatment. The model intercept was set to litter one, as that was when the data collection started. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. Sexes were analyzed separately. A normal distribution was assumed because weight data are continuous. Sample sizes used to assess weight differences include: 45 daughters from 11 cerivastatin cages, 79 daughters from 16 control cages for a total of 124 observations from 27 cages; 48 sons 12 cerivastatin cages and 88 sons from 17 control cages for a total of 136 observations from 29 cages.

Litter size data are discrete counts and therefore were analyzed with a generalized linear mixed model (GLMM) with a Poisson distribution and logarithmic link. This model assessed the effects of treatment, time, and the interaction of time and treatment. The model intercept was set to litter one, as that was when the data collection started. Litter size data were collected at weaning. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. There were 24 cerivastatin-exposed litters from 12 cages and 44 control litters from 18 cages for a total of 68 observations from 30 cages used in this analysis.

Body weight was analyzed with a linear mixed-effects model. This model assessed the effects of treatment, time, and the interaction of time and treatment on the 116 population founders. Sexes were analyzed separately. A normal distribution was assumed because weight data are continuous. Treatment, time, and their interaction were modeled as fixed effects and individual and population were modeled as random effects with random slopes and intercepts generated for each. The intercept was set at week zero, as this was when founders were released into the enclosures and at which collected of weight data from OPAs began. Collection of founder weight data continued on surviving individuals at each pup sweep. There were a total of 438 observations from 76 females and 196 observations from 40 males collected throughout the experiment.

Reproductive outputs were in terms of total offspring and thus are discrete data. These data were analyzed with a generalized linear mixed model with a Poisson distribution and logarithmic link. The model assessed the effects of treatment, time, and the interaction of treatment and time on the population level. These effects were set as

fixed effects in the model and population was set as a random effect with random intercept calculated for each. The intercept was set at week eight, as that was when the first collection period or pup sweep occurred. Reproductive output for each treatment was measured five times over the course of the 28-week study in each of the five independent populations for a total of 50 observations. Female reproductive output was analyzed in terms of total offspring and male reproductive success was analyzed in terms of male offspring.

To assess the probability of territorial ownership, a generalized linear mixed model was used. As a territory can be defended or not, a binomial distribution was used with a logit link. There are six territories within a population and were either occupied by cerivastatin-exposed males, control males, or unoccupied. A total of 130 observations were collected from five populations and analyzed throughout the study. The model assessed the effect of treatment, time, and their interaction. These effects were set as fixed effects and population was set as a random effect with a random intercept generated for each. The model intercept was set to week three, as that was when data existed for each population.

A multivariate Cox proportional hazard model analyzed survival of founders. Impacts of treatment and population were examined in the model. Individuals that survived the length of the study or that were intentionally removed from the study were censored. Female survival was not analyzed because of 100% survival in control females and only three mortalities in the cerivastatin treatment (76 individuals, 3 events). For male survival analyses, 40 individuals were analyzed, 15 events and 25 censorings.

Cox proportion hazard models were conducted in JMP 9.0.3 (SAS institute Inc., Cary NC. Both LMMs and GLMMs were conducted in R 3.0.2 using the lme4 library (Bates *et al*., 2014, R Development Core Team 2013). P-values were calculated for LMM with the Swatterthwaite approximation under the lmerTest function (Kuznetsova *et al.*, 2013). For all mixed models, several candidate models were fit to the data. These models varied in terms of random effects that estimated both intercept and/or slope. For each analysis, the model that explained some of the variance with random effects and had the lowest Akaike Information Criterion (AIC) score was reported. Neither the significance of a fixed effect nor the magnitude of the significance varied between models.

2.11 Supplementary Information References

Bates, D., M. Maechler, B. Bolker, S Walker, R. Christensen, and H. Singmann. 2014. lme4: linear mixed-effects models using S4 classes. R package version:1.1-5.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Kuznetsova, A., Brockhoff, P., and R. Christensen. 2013.lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4 package). R version 2.0-3.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

R Development Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

CHAPTER 3

LOW-DOSE PAROXETINE EXPOSURE CAUSES LIFETIME FITNESS DECLINES IN MALE MICE AS MEASURED BY THE NOVEL ORGANISMAL PERFORMANCE ASSAY

3.1 Abstract

We utilized organismal performance assays (OPAs), a novel toxicity assessment method, to assess the safety of paroxetine. Paroxetine is a selective serotonin reuptake inhibitor (SSRI) that is currently available on the market and is suspected of causing congenital malformations in babies born to mothers who take the drug during the first trimester of pregnancy. OPAs utilize genetically diverse wild mice (*Mus musculus*) to evaluate competitive performance between experimental and control animals as they compete amongst each other for limited resources in seminatural enclosures. Performance measures included reproductive success, male competitive ability, and survivorship. Paroxetine-exposed males weighed 13% less, had 44% fewer offspring, dominated 53% fewer territories, and experienced a 2.5 fold increased trend in mortality when compared with controls. Paroxetine-exposed females had 65% fewer offspring early in the study, but rebounded at later time points. In breeding cages, paroxetine-exposed breeders took 2.3 times longer to produce their first litter and pups of both sexes experienced reduced weight when compared with controls. Low-dose paroxetine-induced health declines

detected in this study were undetected in preclinical trials with doses 2.5-8 times higher than human therapeutic doses. These data indicate that OPAs provide superior sensitivity compared to conventional approaches and demonstrate that OPAs could be a useful tool in safety testing during pharmaceutical development.

3.2 Introduction

Selective serotonin reuptake inhibitors (SSRIs) are the most commonly prescribed antidepressants with 74% of depression cases receiving SSRIs for therapy (Schmidt et al. 1997). Paroxetine [Paxil®, GlaxoSmithKline (GSK), Brentford, England] became commercially available in 1992 and has been used to treat patients with depression, anxiety, and other mood disorders. Paroxetine, along with other SSRIs, are suspected of causing congenital cardiac defects and pulmonary hypertension when a fetus is exposed during the first trimester; however, these epidemiological studies are controversial where some studies find correlative evidence (Williams and Wooltorton 2005, Bérard et al. 2007, Diav-Citrin et al. 2008, Ellfolk and Malm 2010) and others do not (Kulin et al. 1998, Källén and Otterblad Olausson 2007). Despite this controversy, in 2005, the Food and Drug Administration (FDA) requested that paroxetine be labeled as a class D drug (positive evidence of human fetal risk) and issued a warning that paroxetine exposure in the first trimester may potentially cause birth defects (FDA 2005).

The preclinical assessment of paroxetine followed the typical protocol in which the drug was tested to determine whether it was mutagenic, carcinogenic, teratogenic, or if it caused infertility (GSK 2013). No genotoxic effects were detected in rodent cells, and tumors were only detected in mice and rats at doses 2-3.9 times the maximum

recommended human dose (GSK 2013). Teratogenicity was not assessed in mice and no teratogenic effects were in observed in rats when given a dose > 8 times higher than human therapeutic doses (GSK 2013). Paroxetine effects on fertility were not assessed in mice but found to impair rat fertility only at high doses: when females were given 2.5 fold higher and when males were given > 8 times higher than human therapeutic doses (GSK 2013). After successful completion of preclinical assessment, paroxetine was deemed safe and continued onto clinical trials until it was released onto the market.

Like paroxetine, many medications once considered safe are found to cause unacceptable health consequences after public release. On average, 73% of pharmaceuticals fail during clinical trials (Lipsky and Sharp 2001) and 10% of FDA approved pharmaceuticals are recalled after market release (Schuster et al. 2005), despite the 12-15 years of research and $1.4 billion average cost associated with each drug during development (Miller 2012). One cause of the high pharmaceutical failure rate is the inability of current toxicity assessment methods to detect cryptic adversities during preclinical trials, particularly those present at low doses and/or those occurring at low incidences.

We have developed a novel toxicity assessment research method that may be useful during preclinical assessment, known as the organismal performance assay (OPA). In several instances, OPAs have proven capable of detecting mammalian health declines with higher sensitivity than current *in vivo* and *in vitro* approaches. OPAs utilize genetically diverse wild-derived mice (*Mus musculus*) that compete amongst each other for limited resources in seminatural enclosures, which allows for direct competition between treatment and control individuals. Individual performance is measured in terms

of Darwinian fitness (i.e., reproductive success) and components leading to fitness (i.e., survivorship and male competitive ability). OPAs are sensitive because they challenge most physiological systems synergistically and in order for individuals to be successful, high performance from these physiological systems is required.

OPAs have previously been used to quantify the adverse effects of sibling-level and cousin-level inbreeding (Meagher et al. 2000, Ilmonen et al. 2008), harboring a selfish gene (Carroll et al. 2004), and recently, they were the first assay to reveal the adverse effects of added sugar consumption at human-relevant levels (Ruff et al. 2013). In all of these studies, OPAs found substantial deleterious effects that were missed by current methodologies. OPAs contain an internal control (direct competition between treatment and control animals), which prevents OPAs from generating false positives.

Here OPAs are used to determine if paroxetine exposure near human therapeutic doses during *in utero* and into early adulthood cause fitness declines in wild mice. Our findings address the current lack of information concerning long-term paroxetine exposure on health at the organismal level. If paroxetine exposure adversely affects any physiological system, we hypothesize that exposed individuals will suffer survival and reproduction declines relative to control individuals while competing within OPAs. Furthermore, while generating animals for OPAs, we assessed whether paroxetine exposure negatively affects reproduction of exposed breeders and the weight of the resulting offspring.

3.3 Materials and Methods

3.3.1 Animals. Wild-derived outbred house mice were used in this experiment. Unlike many genetically inbred mouse strains, wild mice have behavioral characteristics that allow them to function in natural and seminatural environments (Nelson et al. 2013). In this experiment, individuals were from the 12th generation of the colony described by Meagher et al. 2000. Genetic diversity of this colony was assessed in the 11th generation and found to be comparable to wild populations (Cunningham et al. 2013). Within OPAs and breeding cages, individuals were provided access to food and water *ad libitum* and maintained on a 12:12 hour light:dark cycle. All procedures were approved by the University of Utah IACUC.

3.3.2 Drug exposure. Dosing was achieved by incorporating 7.5 g paroxetine (GSK, molecular formula: $C_{19}H_{20}FNO_3 \cdot HCl$) into 50 kg of rodent chow (TD.130006; Harlan Teklad, Madison, WI). Mice consuming an average of 3 g of food per day and weigh 20 g will ingest 0.45 mg per day or 22.5 mg/kg/day. Using a standard metabolic rate conversion factor, this is equivalent to a human dose of 1.82 mg/kg/day, or a daily dose of 109.20 mg, assuming the average human weighs 60 kg (Reagan-Shaw et al. 2008). Given that paroxetine is prescribed in the range of 20 – 60 mg/day (Dunner and Dunbar 1992, GSK 2013), our dose is 1.82 fold higher than human therapeutic doses, yet lower than doses used in previous animal studies (Coleman et al. 1999, Rayburn et al. 2000, El-gaafarawi et al. 2005). One study determined that a paroxetine dose of 30 mg/kg/day achieved serum levels in mice that were comparable to human serum levels when taking the highest therapeutic dose (Coleman et al. 1999).

Sixty breeder pairs were selected for this experiment; 20 pairs were exposed to paroxetine while the remainder served as controls. The asymmetry in cage number is due to the production of additional control animals for another study. Prior to breeding, animals were individually housed and provided with their respective diets. Females were exposed to paroxetine eight days prior and males five days prior to breeding. Breeding pairs were kept together until a maximum of four litters were produced to ensure enough animals for OPA assessment. At 28 days of age, pups were weaned and housed in same-sex sibling cages. Upon weaning, individual weight, sex, and litter size data were collected and paroxetine exposure continued until animals were released into enclosures. The duration of paroxetine exposure maximized the ability of OPAs to detect health consequences because once animals were released into the seminatural enclosures, they were all fed the control diet as currently, we are unable to keep animals on their respective diets while they are free ranging during OPAs. Upon release into OPAs, both paroxetine-exposed and control animals were provided with the control diet *ad libitum.* Switching the paroxetine-exposed animals to the control diet was the most conservative approach of detecting fitness impacts, because OPAs would then be assessing cumulative damage.

3.3.3 OPA enclosures. Enclosures have previously been described in Ruff et al. 2013. Briefly, the indoor enclosures are approximately 30 m^2 and consist of two types of territories, optimal (n = 4) and suboptimal (n = 2). Each optimal territory contained a defendable box with multiple dark nesting sites and direct access to food. Suboptimal territories contained two nesting boxes exposed to light and had indirect access to food.

Territories were separated by hardware mesh that is easily climbed, but added an element of spatial complexity (Figure S3.1).

Five independent OPA populations were established and maintained for 28 weeks. OPA populations consisted of eight males and 14-16 females, for a total of 116 animals (40 males, 76 females); these animals are referred to as population founders. Half of the individuals of each sex were paroxetine-exposed while the remainder served as controls; this population structure allows paroxetine individuals to directly compete with control individuals for mates, resources, and territories. Enclosure space and population size created a population density within the range observed in the wild (Sage 1981).

Upon OPA release, male mice were on average 14.3 (SD ± 4.7) weeks old and females were 18.9 (± 7.0) weeks old. Males were released into enclosures with non-experimental females to allow males to establish territories, which prevented incidental breeding during the initial social chaos of population formation. Nonexperimental females were then removed and replaced with experimental females one week later. Male population founders were unrelated at the cousin level or above with the exception that three populations consisted of one set of brothers and one population consisted of two sets of brothers. Similarly, all populations consisted of sister pairs and four populations consisted of sister trios. At most, one male per treatment was in a population containing two of his sisters. When relatedness was present, it was balanced between treatments.

3.3.4 Reproductive success. Reproductive success of founders was determined by removing and genotyping all offspring in OPAs. Every five weeks, research personnel conducted a "pup sweep," in which all offspring were removed, sacrificed, and had a tissue sample collected for genetic analyses. As the gestation period of mice is three

weeks, the first pup sweep occurred on week eight, then every sweep following occurred on five-week intervals; this sweep schedule prevented offspring from reaching sexual maturity within OPAs. A total of 872 samples were collected with an average of 174.4 ± 38.1 (M ± S.D.) offspring per population.

In three of the five populations, reproductive success was determined on a population-level by examining sex-specific allelic variants that have been previously described in Meagher et al. 2000. Briefly, nonoverlapping allelic variants were assigned to founders of each treatment (paroxetine-exposed or control). Females were selected upon allelic variants of the mitochondrial genome and males upon the Y-chromosome. The output of male reproductive success in these populations is based solely upon male offspring, as females do not possess Y-chromosomes. Reciprocal markers were assigned across populations to control for confounding effects. Mitochondrial genotypes were assessed in 626 samples (three of five populations) and obtained for 100% of offspring. Y-chromosome genotypes were assessed in all five populations to determine male reproductive output and the sex ratio of offspring. Of the 872 offspring, 414 Y-chromosome genotypes were obtained suggesting that 95% of all males were typed if the sex ratio was 1:1; successfully genotyping nearly all of the progeny allowed for a thorough analysis of founder reproductive success.

To gain a better understanding of individual-level reproductive success, parentage analysis was conducted in two populations using multiple microsatellite loci. Female reproductive success determined by microsatellite loci were converted to population level readouts (i.e., number of pups per treatment rather than number of pups per individual) and combined with the mitochondrial data for analysis. Between six and 17 autosomal

microsatellite loci were amplified, scored, and analyzed in a stepwise fashion. Loci used were: d1mit251, d1mit449, d3mit22, d3mit312, d3mit333, d4mit205, d5mit139, d6mit138, d9mit232, d9mit251, d12mit277, d14mit128, d17mit24, d17mit62, d17mit82, d17mit102, and d19mit110. Primer sequences were obtained from the Mouse Genome Informatics website, The Jackson Laboratory, Bar Harbor Maine (http://www.informatics.jax.org/ accessed March 2014). Primers were tagged with either CY-5 or CY-3 fluorescent dye. DNA samples were PCR-amplified and then run on 14" x 17", 6.25% denaturing acrylamide gel at 40 W for three to seven hours (locus dependent). Gels were imaged on a Typhoon Scanner 8600 using ImageQuant software (Amersham Biosciences, Piscataway, NJ).

Parentage via multiple microsatellites was assigned using Cervus 3.0 (Kalinowski et al. 2007). Genotypes of all candidate mothers and fathers and all offspring within each population were used to calculate allele frequencies. Simulations were run 10,000 cycles with an error rate of 1% to derive a delta score. Assigned parents were accepted when the trio confidence of mother, father and offspring was 95%. Using this rule, 91% (187/205) of one population was genotyped and 75% (147/195) of the second population was genotyped.

3.3.5 Male competitive ability. For identification purposes, all animals were given a unique ear punch and received a passive integrated transponder (PIT) tag (TX1400ST, BioMark, Boise, ID) prior to OPA release. Two sets of PIT tag antenna and readers (FS2001F-ISO, BioMark, Boise, ID) collected data on OPA populations and were rotated twice weekly among concurrent populations. Antennas were placed on top of feeding stations in each territory of an enclosure. PIT tag data were downloaded to a computer

containing data logging software (Minimon, Culver City, CA). A male was assigned as an territorial occupant when he possessed > 80% of the total reads at a particular location. PIT tag data were collected on female mice, but were not analyzed due to the general lack of information on female dominance behavior.

3.3.6 Survivorship. Noninvasive health checks were performed daily and extensive enclosure checks every five weeks during pup sweeps. Extensive enclosure checks were limited as to avoid disrupting territoriality that increases the rate of infanticide. When deceased founders were observed, they were removed from the enclosures. The date of death was estimated upon the condition of the corpse. Severely decomposed founders were given a date half way between the date the individual was found and the last date the PIT tag of that individual was recorded.

3.3.7 Statistical analyses. Cox proportional hazard (PH) models were used for OPA survivorship and the time to produce first litter from cage data (JMP 9.0.3, SAS institute Inc., Cary NC). Generalized linear mixed models (GLMM) were used for litter size, reproductive success, and male competitive ability (R 3.0.2). Reported SEMs are asymmetric because values have been back-transformed from logarithmic data. Linear mixed models (LMM) were used for wean weight and OPA body weight (R 3.0.2). A complete description of statistical analyses can be found in the supplementary information.

3.4 Results

3.4.1 Breeding cage measures. Paroxetine-exposed breeders took 2.3 times longer to produce their first litter when compared to controls (PH; $\chi^2 = 3.98$, $p < 0.05$; Figure

3.1); however, litter size was not affected by treatment (GLMM; $z = -0.58$, $p = 0.56$). Control breeders produced an average of 4.54 (SEM +0.45, −0.41) pups in their first litter, while paroxetine-exposed breeders produced 4.09 (+0.83, −0.75) pups in their first litter. No effect of time (GLMM; $z = -0.70$, $p = 0.49$) or time by treatment was detected (GLMM; $z = 0.65$, $p = 0.52$). For a complete readout of mixed model results for breeding cage data, see Table S3.1.

Offspring from paroxetine-exposed breeders weighed less at weaning than offspring from control breeders. Paroxetine-exposed female offspring weighed 16% less than controls with an average of 8.81 g (SEM ± 0.52), whereas control female offspring weighed 10.77 g (± 0.31) (LMM; $t = -3.70$, $p < 0.001$; Figure 3.2A). No effect of parity on weight occurred (LMM; $t = 1.80$, $p = 0.11$), nor was there an interaction between treatment and parity (LMM; $t = 0.43$, $p = 0.68$). Male offspring from paroxetine-exposed breeders weighed 25% less than controls (LMM; $t = -3.83$, $p < 0.001$; Figure 3.2B), with pups weighing 9.63 g (± 0.73) and control male offspring weighing 12.39 g (± 0.37). No effect of parity (LMM; $t = 1.34$, $p = 0.18$) or parity by treatment was detected (LMM; $t = 0.41$, $p = 0.69$).

3.4.2 OPA measures. Within OPAs, paroxetine exposure had differential affects on body weight between the sexes. Upon OPA release (model intercept, week zero), no differences were detected in body weight of paroxetine-exposed and control females (LMM; $t = -1.04$, $p = 0.30$; Figure 3.3A). Paroxetine-exposed females weighed on average 19.39 g (SEM ± 1.03) while controls weighed 20.46 g (± 1.25). Both groups of females gained weight over time (LMM; $t = 5.12$, $p < 0.01$), due to pregnancy, and no interaction of time by treatment was detected (LMM; $t = -1.64$, $p = 0.11$). Paroxetine-

exposed males weighed less (LMM; $t = -3.94$, $p < 0.001$; Figure 3.3B) than controls upon OPA entrance. Paroxetine-exposed males weighed 18.42 g (± 0.87) and controls weighed 21.75 g (± 0.90). Both groups gained weight over time (LMM; $t = 2.60$, $p = 0.05$), but no time by diet interaction was detected (LMM; $t = -0.22$, $p = 0.83$), indicating that paroxetine-exposed males weighed 13% less than controls throughout the duration of the experiment. For a complete readout of mixed model results for OPA body weight measures, see Table S3.2.

Female reproductive success was hindered by paroxetine exposure (GLMM; $z = -5.03$, $p < 0.0001$; Figure 3.4A) at week eight (model intercept), where mean reproduction of paroxetine-exposed females was 65% less than controls. Paroxetine-exposed females had an average of 10.68 (+1.44, −1.26) offspring per population, while control females had an average of 20.16 (+1.63, −1.51) offspring per population. No effect of time was detected (GLMM; $z = -0.63$, $p = 0.53$). However, there was a significant interaction between time and treatment (GLMM; $z = 3.95$, $p < 0.0001$), suggesting that paroxetine-exposed female reproduction increased over time. As exposure had significant and opposing effects in regards to the intercept and slope of the linear model, post-hoc t tests were conducted at each pup sweep and a significant difference was only detected at week eight ($p < 0.05$). For a complete readout of mixed model results for OPA reproduction and competitive ability, see Table S3.3.

Paroxetine exposure also negatively affected male reproductive success as measured by male offspring, where paroxetine-exposed males had 44% fewer offspring than controls (GLMM; $z = -2.72$, $p < 0.01$; Figure 3.4B). At week eight (model intercept), mean reproduction of paroxetine-exposed males was on average 5.25 (SEM

+1.07, −0.90) male offspring per population per population. Control males sired an average of 8.71 (+1.09, −0.96) male offspring per population. Male founders from both treatments had an increase in reproductive success over time (GLMM; $z = 2.16$, $p = 0.03$), but no time by treatment interaction was detected (GLMM; $z = -0.51$, $p = 0.61$), indicating that paroxetine-exposed males had reduced reproductive success throughout the study.

Male competitive ability was adversely impacted by paroxetine exposure, where paroxetine males occupied 53% fewer territories than controls. At week three (model intercept) control males occupied 47% of the territories, while paroxetine-exposed males only occupied 22% territories (GLMM; $z = -4.11$, $p < 0.0001$; Figure 3.5), leaving 31% of the territories undefended. The percent of undefended territories is not unusual because 2/6 (or 33%) were suboptimal and often difficult to defend. There was no difference in territorial occupancy over time (GLMM; $z = -0.14$, $p = 0.89$), nor was there a difference in the time by diet interaction (GLMM; $z = 0.18$, $p = 0.86$), indicating that the differential acquisition of territories by control and paroxetine-exposed males was consistent across the study.

No differences were detected in mortality between paroxetine-exposed and control females (PH; $\chi^2 = 0.66$, $p = 0.42$; Figure 3.6A). Female mortality rates did not differ in replicate populations (PH; $\chi^2 = 3.51$, $p = 0.48$), nor was there a difference in the effect of treatment among populations (PH; $\chi^2 = 3.35$, $p = 0.50$). A marginally statically significant trend was detected in which male mortality was increased approximately 2.5 by paroxetine exposure (PH; $\chi^2 = 3.27$, $p = 0.07$; Figure 3.6B). Male mortality rate did

not differ in replicate populations (PH; $\chi^2 = 5.24$, $p = 0.26$), nor was there a difference in the effect of treatment among populations (PH; $\chi^2 = 3.77$; $p = 0.44$).

3.5 Discussion

We detected numerous adverse effects when animals were exposed to paroxetine *in utero* through early adulthood. In cages, breeders took 2.3 times longer to produce their first litter and offspring from these breeders experienced reduced wean weight when compared with controls. Additionally, OPAs revealed several deleterious effects in paroxetine-exposed males in which they experienced 13% reduced body weight, had 44% reduced reproductive success, occupied 53% fewer territories, and showed a marginally significant 2.5-increased mortality rate. Paroxetine-exposed females too suffered reduced reproductive success by 65%, but only at the initial pup sweep, suggesting they might be recovering from their paroxetine exposure. OPAs revealed numerous health consequences at the organismal level, which were missed by conventional methods; these phenotypes are likely to be caused by paroxetine disrupting several mechanisms, whose descriptions follow.

Paroxetine-exposed breeders took 2.3 times longer to produce their first litter when compared with controls, which could be caused by paroxetine interfering with hormones that regulate the female reproductive system. Within the estrus cycle, a surge of luteinizing hormone (LH) occurs at the time of ovulation allowing for fertilization of the ovum. A reduction in LH levels, along with other important reproductive hormones (i.e., estradiol, progesterone) have been observed after paroxetine exposure in female rats (El-gaafarawi et al. 2005). These hormones play an important role in regulating the

function of the female reproductive system and alterations of these hormones may have caused delayed reproduction.

Although litter size did not differ between treatments in breeding cages, paroxetine-exposed founder females had 65% fewer offspring, but only at the first pup sweep in OPAs. This finding suggests that OPAs are more sensitive than cages as they were able to detect this fitness consequence, which was invisible under standard housing. It is likely that the damage from previous paroxetine exposure decreased over time as exposure ceased at the time of OPA release, which may in part explain the increase in reproductive success of exposed females as the study progressed. However, the initial reduction in reproduction may also be explained by the endocrine disruptions responsible for the delayed reproduction experienced by females in cages. The data indicating that the negative effects of paroxetine exposure on reproduction decrease over time suggest that the negative effects on female reproduction is reversible.

Wean weight was reduced in both male and female offspring when breeding pairs were exposed to paroxetine. Reduced birth weight has been detected in previous rodent (Coleman et al. 1999, Rayburn et al. 2000) and human studies (Diav-Citrin et al. 2008). Upon OPA entrance, the affect of treatment on female weight was no longer detected, although a trend was observed wherein paroxetine-exposed females gained weight at a decreased rate relative to controls. This trend of reduced weight within OPAs may be explained by reduced pregnancies in the paroxetine-exposed females. Like females, paroxetine-exposed males weighed less at weaning and continued to weigh less throughout all time points within the OPAs.

Paroxetine-exposed males were less competitive and dominated 53% fewer territories than controls. Decreased body weight may be a partial cause of the decreased competitive ability, as males must engage in physical competitions to obtain and defend territories. The benefit of a relatively larger body weight on competitive ability is well established to be advantageous in territorial acquisition and defense in mice [e.g., (van Zegeren 1980, Krackow 1993)]. Another possible explanation of why paroxetine-exposed males are less competitive is that the drug may have altered their hormonal signaling. Testosterone plays a large role in male behavior, especially competitive behaviors in mice (Zielinski and Vandenbergh 1993). El-Gaafarawi et al. (2005) found that when male rats were exposed to paroxetine, serum levels of testosterone were significantly reduced and estradiol levels were increased. Although a proximate mechanistic cause of paroxetine is suggested here, the adverse effects on organismal competition have not previously been demonstrated. It is likely that multiple factors contribute to these changes including hormonal alterations that may have changed normal behaviors, such as the desire to participate in intrasexual competition.

Paroxetine-exposed males experienced reduced reproductive success, producing 44% fewer offspring than controls. Previous OPA studies reveal that dominant males sire > 80% of offspring within the enclosures (Carroll et al. 2004) and therefore, differences in competitive ability likely explain a large portion of the differential reproduction. A trend was also observed in which paroxetine-exposed males suffered greater mortality and thus even if not significant alone, it does contribute to the reduced reproduction. Another possibility is that paroxetine has direct negative impacts on the male reproductive system. Paroxetine exposure has been shown to cause reduced sperm count

(Baldwin et al. 1989, El-gaafarawi et al. 2005) and an increased rate of sperm abnormalities in rats (El-gaafarawi et al. 2005). Altered hormone levels such as decreased follicle stimulating hormone (FSH), LH, and testosterone have been documented in paroxetine-exposed rats (El-gaafarawi et al. 2005) and these hormones are important in spermatogenesis (Kovacs 2012). Decreased competitive ability, increased mortality, and insults to sperm production are likely causes that lead to a decline in male reproductive success.

Although fitness declines were detected in both sexes, male mice exposed to paroxetine suffered greater adversity than did females. Differential health consequences between sexes have been detected with the use of OPAs in previous studies. Males have suffered greater negative fitness impacts than females in three previous OPA studies, inbreeding at the cousin level (Ilmonen et al. 2008), inbreeding at the sibling level (Meagher et al. 2000), and when animals were fed a moderate sugar diet (Ruff et al. 2013; Figure S3.2). Male mice may have been adversely affected by the paroxetine treatment more so than females due to the competitive nature of their social ecology. Since OPA studies generate the same outputs, relative fitness can be compared between treatments and for example, results from this study indicate that paroxetine-exposed male fitness (i.e., reproductive success) is ~25% less than the inbred males born to parents of first cousins (Figure S3.2).

3.6 Conclusions

Despite the controversial evidence of paroxetine-induced birth defects in humans, we detected health consequences in mice that underwent low-dose paroxetine exposure *in*

utero and into early adulthood. While paroxetine-exposed females recovered from the deleterious effects after the exposure ceased but males continued to experience severe adverse health consequences throughout their lifetime. Here, OPAs indicate that paroxetine exposure causes adversity without ambiguous interpretation and we can conclude that paroxetine exposure *in utero* and into early adulthood negatively affects the health of the resulting offspring, particularly male offspring. These risks should be considered while deciding whether or not to take paroxetine during pregnancy.

There is currently a great need for a more sensitive toxicity assessment assay that provides results without ambiguous interpretation of adversity. Data presented here suggest OPAs will fulfill that role and will be useful during pharmaceutical development and if implemented have the potential to significantly reduce the failure rate during clinical trials. Likewise, OPAs may also be useful in two other divisions of toxicity assessment: 1) screening for possible environmental pollutants before humans and wildlife undergo long-term exposures under the assumption that they are safe and 2) assessing the health consequences of dietary components. Undeniably, the revelation that low dose paroxetine-exposure negatively affected weight, competitive ability, reproduction, and survival is evidence that OPAs have the potential to identify and quantify toxicities that have escaped detection in previously used systems. These organismal phenotypes can now form the basis for investigations to discover the underlying mechanisms.

3.7 Acknowledgments

We thank D. Tripodi for suggesting that we apply OPAs to assess pharmaceuticals; M. Bartlett, S. Eddy, D. Kircher, E. Schwab, M. Sosa, P. Wheatley, and C. Young for data collection and genotyping; A. Bwika and J. Gale for animal care. The project was funded by the University of Utah's Technology Commercialization Program and was partially conducted while W.K.P was supported by NSF grant DEB 09-18969 and NIH grant R01-GM109500. S.M.G. was supported by an NSF GK-12 Educational Outreach Fellowship (DGE 08-41233).

3.8 References

Baldwin, J., E. Davidson, A. Pritchard, and J. Ridings. 1989. The reproductive toxicology of paroxetine. Acta Psychiatrica Scandinavica 80:37–9.

Bérard, A., E. Ramos, E. Rey, L. Blais, M. St-André, and D. Oraichi. 2007. First trimester exposure to paroxetine and risk of cardiac malformations in infants: the importance of dosage. Birth Defects Research. Part B, Developmental and reproductive toxicolocy 80:18–27.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Coleman, F. H., H. D. Christensen, C. L. Gonzalez, and W. F. Rayburn. 1999. Behavioral changes in developing mice after prenatal exposure to paroxetine (Paxil). American journal of obstetrics and gynecology 181:1166–71.

Cunningham, C. B., J. S. Ruff, K. Chase, W. K. Potts, and D. R. Carrier. 2013. Competitive ability in male house mice (Mus musculus): genetic influences. Behavior genetics 43:151–60.

Diav-Citrin, O., S. Shechtman, D. Weinbaum, R. Wajnberg, M. Avgil, E. Di Gianantonio, M. Clementi, C. Weber-Schoendorfer, C. Schaefer, and A. Ornoy. 2008. Paroxetine and fluoxetine in pregnancy: a prospective, multicentre, controlled, observational study. British journal of clinical pharmacology 66:695–705.

Dunner, D. L., and G. C. Dunbar. 1992. Optimal dose regimen for paroxetine. Journal of Clinical Psychiatry 53:21–6.

El-gaafarawi, I., M. Hassan, G. Fouad, and F. El-komey. 2005. Toxic effects of paroxetine on sexual and reproductive functions of rats. The Egyptian Journal of Hospital Medicine 21:16–32.

Ellfolk, M., and H. Malm. 2010. Risks associated with in utero and lactation exposure to selective serotonin reuptake inhibitors (SSRIs). Reproductive toxicology 30:249–60.

FDA. 2005. Public Health Advisory: Paroxetine. Accessed March 2014. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/PublicHealthAdvisories/ucm051731.htmhttp://www.fda.gov.

GSK. 2013. Paxil (paroxetine hydrochloride) Tablets and Oral Suspension. Accessed March 2014. http://www.apotex.com/us/en/products/downloads/pil/paxil_irtb_ins.pdf.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular ecology 16:1099–106.

Källén, B. A. J., and P. Otterblad Olausson. 2007. Maternal use of selective serotonin re-uptake inhibitors in early pregnancy and infant congenital malformations. Birth defects research. Part A, Clinical and molecular teratology 79:301–8.

Kovacs, W. 2012. Textbook of Endocrine Physiology. Pages 239–263 (W. Kovacs and S. Ojeda, Eds.). Sixth edition. Oxford University Press, Inc, New York.

Krackow, S. 1993. The effect of weaning weight on offspring fitness in wild house mice (Mus musculus domesticus): A preliminary study. Ethology 95:76–82.

Kulin, N. a, A. Pastuszak, S. R. Sage, B. Schick-Boschetto, G. Spivey, M. Feldkamp, K. Ormond, D. Matsui, a K. Stein-Schechman, L. Cook, J. Brochu, M. Rieder, and G. Koren. 1998. Pregnancy outcome following maternal use of the new selective serotonin reuptake inhibitors: a prospective controlled multicenter study. The journal of the American Medical Association 279:609–10.

Lipsky, M. S., and L. K. Sharp. 2001. From idea to market: the drug approval process. The Journal of the American Board of Family Practice 14:362–7.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

Miller, H. 2012. How is the FDA really doing? Genetic Engineering and Biotechnology News 32:6–8.

Nelson, A. C., J. W. Cauceglia, S. D. Merkley, N. A. Youngson, A. J. Oler, R. J. Nelson, B. R. Cairns, E. Whitelaw, and W. K. Potts. 2013. Reintroducing domesticated wild mice to sociality induces adaptive transgenerational effects on MUP expression. Proceedings of the National Academy of Sciences 110:19848–53.

Rayburn, W. F., C. L. Gonzalez, H. D. Christensen, T. C. Kupiec, J. A. Jacobsen, and J. D. Stewart. 2000. Effect of antenatal exposure to paroxetine (Paxil) on growth and physical maturation of mice offspring. The Journal of Maternal-Fetal Medicine 9:136–141.

Reagan-Shaw, S., M. Nihal, and N. Ahmad. 2008. Dose translation from animal to human studies revisited. The Federation of American Societies for Experimental Biology Journal 22:659–61.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

Sage, R. 1981. The Mouse in Biomedical Research. Pages 40–90 (H. Foster, J. Small, and J. Fox, Eds.). Academic Press, New York.

Schmidt, L., B. Greenberg, Holzman G, and J. Schulkin. 1997. Treatment of depression by obstetrician-gynecologists: a survey study. Obstetrics and Gynecology 90:296–300.

Schuster, D., C. Lagger, and T. Langer. 2005. Why drug fail - A study on side effects in new chemical entities. Current Pharmaceutical Design 11:3545-59.

The Jackson Laboratory. 2013. Accessed March 2014. http://www.informatics.jax.org.

Williams, M., and E. Wooltorton. 2005. Paroxetine (Paxil) and congenital malformations. Canadian Medical Association Journal 173:1320–1.

Van Zegeren, K. 1980. Variation in aggressiveness and the regulation of numbers in house mouse population. Netherlands Journal of Zoology 30:635–770.

Zielinski, W. J., and J. G. Vandenbergh. 1993. Testosterone and competitive ability in male house mice, Mus musculus: laboratory and field studies. Animal Behaviour 45:873–91.

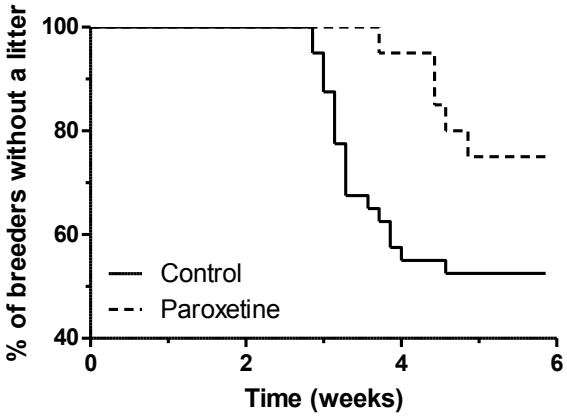

Figure 3.1. Time until first litter of paroxetine-exposed and control breeders. In breeding cages, paroxetine-exposed breeders (n = 20) took 2.3 times longer to produce their first litter when compared with control breeders [n = 40, (PH; $\chi^2 = 3.98$, $p < 0.05$)].

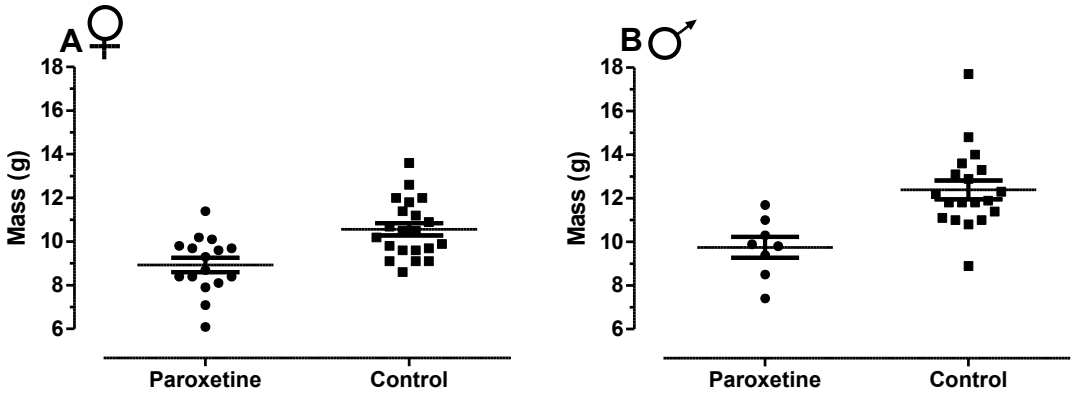

Figure 3.2. Offspring wean weight from paroxetine-exposed and control breeders. **A)** Female offspring from paroxetine-exposed breeders weighed 16% less than controls [n = 24 cages, *obs* = 134 (LMM; t = −3.70, p < 0.001)]. **B)** Male offspring from paroxetine-exposed breeders weighed 25% less than male offspring from control breeders [n = 25 cages, *obs* = 111 (LMM; t = −3.83, p < 0.001)]. Lines represent means and error bars, standard error.

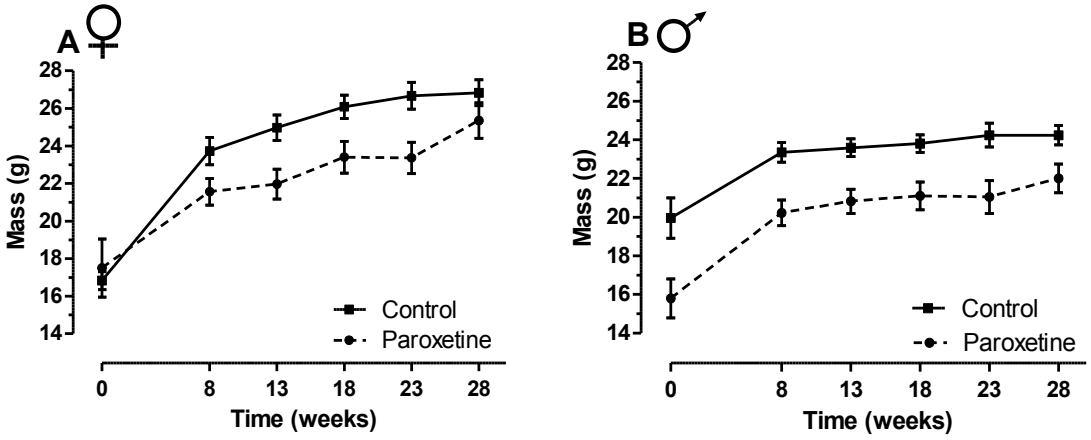

Figure 3.3. Body weight of paroxetine-exposed and control founders within OPAs over time. A) Female body weight did not differ between treatments [n = 76 mice, *obs* = 386 (LMM; t = −1.04, p = 0.30)]. **B)** Paroxetine-exposed males weighed 13% less than controls over the duration of the study [n = 40 mice, *obs* = 179 (LMM; t = −3.94, p < 0.001)]. Lines connect means and error bars represent standard error.

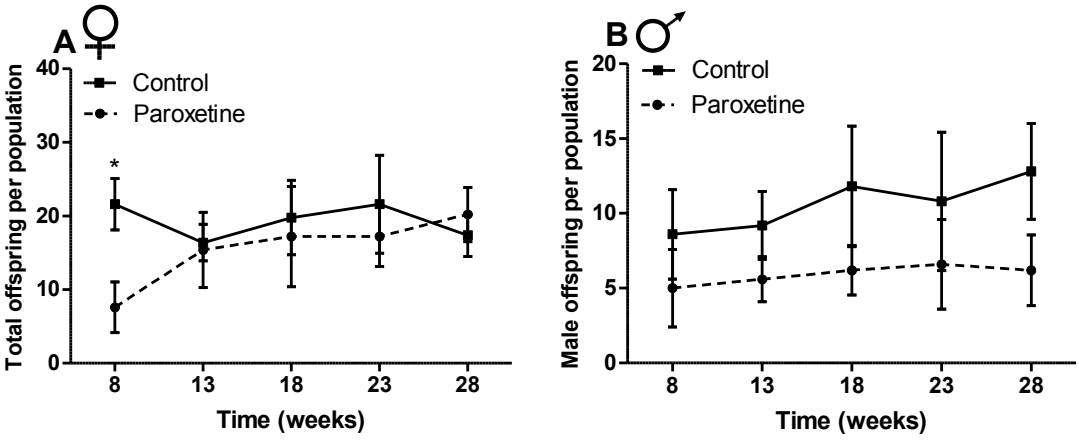

Figure 3.4. Reproductive success of paroxetine-exposed and control founders in OPAs. A) Paroxetine-exposed females had 65% fewer offspring at week eight [n = 5 populations, *obs* = 50 (GLMM; z = −5.03, p < 0.0001)], but had more offspring over time (GLMM; z = 3.95, p < 0.0001). This was confirmed with post-hoc tests indicating that at week eight, reproduction was significantly different; however, no difference was seen at any other time. **B)** Paroxetine-exposed males had 44% fewer male offspring compared with controls across the study [n = 5 populations, *obs* = 50 (GLMM; z = −2.72, p < 0.01). Lines connect means of the five populations at each time point for each sex and error bars represent standard error.

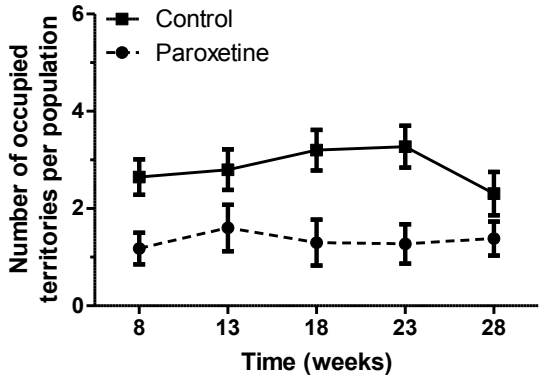

Figure 3.5. Competitive ability of paroxetine-exposed and control males over time within OPAs. Paroxetine-exposed males occupied 54% fewer territories than controls over the duration of the study [n = 5 populations, obs = 122 (GLMM; $z = -4.11$, $p < 0.0001$). Control males occupied 47% of the territories paroxetine-exposed males occupied 22% territories and the remaining 31% of the territories were undefended. A male was considered a territorial occupant if > 80% of his reads were at a particular location. Points represent the number of territories occupied by males from each treatment within five populations over a multiday reading frame. Lines represent the best fit of the data. Paired observations (one paroxetine point for every control) occur at each reader session. When data points from multiple populations overlapped, x-axis scatter was created for visual clarity.

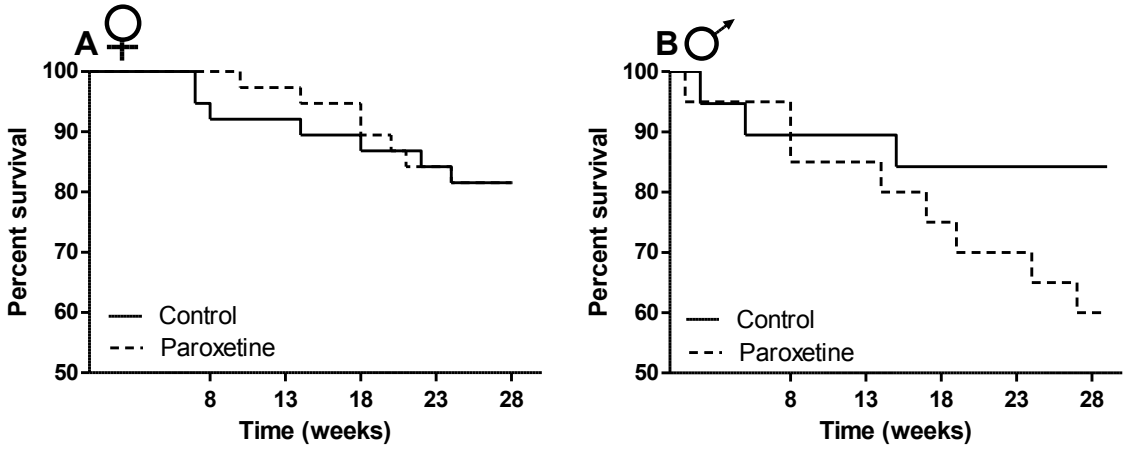

Figure 3.6. Survivorship of paroxetine-exposed and control animals within the OPAs. A) No differences in mortality were detected between treatments in females [n = 76 (PH; $\chi^2 = 0.66$, p = 0.42)] **B)** Paroxetine-exposed males (n = 40) experienced a marginally significant, twofold increase in mortality rate when compared to controls (PH; $\chi^2 = 3.27$, p = 0.07).

3.9 Supplementary Information

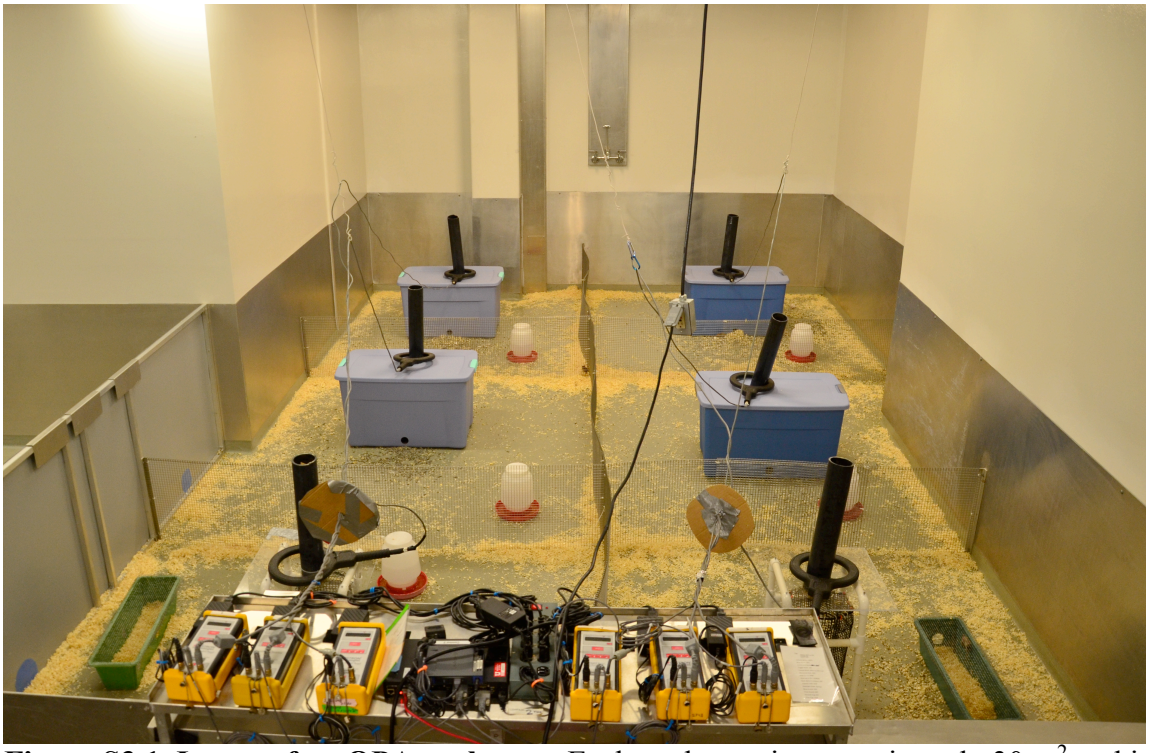

Figure S3.1. Image of an OPA enclosure. Each enclosure is approximately 30 m^2 and is divided into six territories by hardware mesh. The four territories containing the large blue storage bins are optimal, which are defendable and contain multiple dark nesting sites. The suboptimal territories contain light exposed nesting boxes (plantar boxes with wire tops). Each territory contains its own water (poultry waterers) and access to food (black chimneys). PIT tag antennae (tennis racket like structures) are placed atop the feeding stations. Photograph courtesy Doug Cornwall.

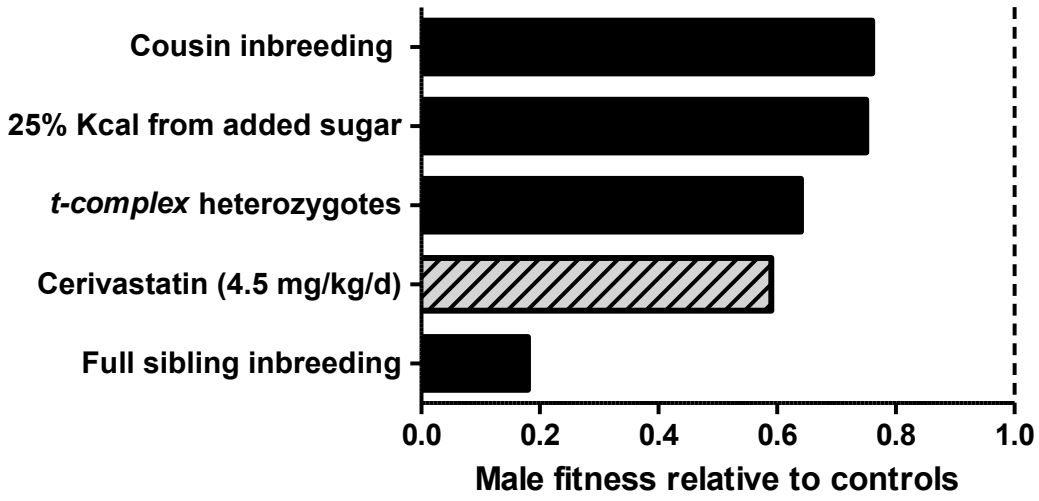

Figure S3.2. Comparison of male reproductive success relative to control counterpart from published OPA experiments. Paroxetine exposure had greater consequences on male fitness (reproductive success) when compared with cousin inbreeding, moderate sugar consumption, and a selfish gene, but suffered less fitness consequences when compared to inbreeding at the sibling level (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013). Figure modified from Ruff et al. 2013.

Table S3.1. Mixed model results for litter size and wean weight between paroxetine-exposed and control breeders.

Litter Size	GLMM with Poisson distribution and logarithmic link (groups = 28; observations = 70)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.0207	0.1441		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Litter 1)	1.5125	0.0941	16.0760	<0.0001***
Treatment (Paroxetine)	-0.1128	0.1945	-0.5800	0.5620
Parity	-0.0533	0.0762	-0.6990	0.4850
Treatment (Paroxetine)×Parity	0.1221	0.1879	0.6500	0.5160
Female Wean Weight	LMM (groups = 24; observations = 134)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.7813	0.8839		
Cage (Slope)	0.2334	0.4831		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Litter 1)	10.7732	0.3143	34.2770	<0.0001***
Treatment (Paroxetine)	-1.9583	0.5286	-3.7050	0.0009***
Parity	0.3930	0.2188	1.7960	0.1074
Treatment (Paroxetine)×Parity	0.1773	0.4137	0.4290	0.6765
Male Wean Weight	LMM (groups = 25, observations = 111)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept First Litter)	1.0417	1.0207		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Litter 1)	12.3919	0.3750	33.0470	<0.0001***
Treatment (Paroxetine)	-2.7664	0.7215	-3.8340	0.0003***
Parity	0.2289	0.1713	1.3360	0.1843
Treatment (Paroxetine)×Parity	0.1867	0.4591	0.4070	0.6852

*** Indicates a p value <0.001.

Table S3.2. Linear mixed model results for body weight over time within OPAs.

Female Body Weight	LMM (groups = 76, observations = 386)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	10.3448	3.2163		
Individual (Slope)	0.0134	0.1157		
Population (Intercept)	5.0655	2.2507		
Population (Slope)	0.0079	0.0887		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Week 0)	20.4628	1.2548	16.3070	<0.0001***
Treatment (Paroxetine)	-1.0752	1.0308	-1.0430	0.3005
Time	0.2731	0.0533	5.1200	0.0026**
Treatment (Paroxetine)×Time	-0.0807	0.0493	-1.6360	0.1064
Male Body Weight	LMM (groups = 40, observations = 179)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	5.6050	2.3675		
Population (Intercept)	3.1113	1.7639		
Population (Slope)	0.0069	0.0831		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Week 0)	21.7898	0.9893	22.0250	<0.0001***
Treatment (Paroxetine)	-3.3184	0.8415	-3.9430	0.0003***
Time	0.1049	0.0404	2.5950	0.0536
Treatment (Paroxetine)×Time	-0.0050	0.0232	-0.2150	0.8303

** Indicates a p value < 0.01, *** < 0.001.

Table S3.3. Generalized linear mixed model results for reproduction and male competitive ability over time within OPAs.

Female Reproduction	GLMM with Poisson distribution and logarithmic link (groups = 5; observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Slope)	0.0004	0.0175		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 8)	3.0038	0.0778	38.6300	<0.0001***
Treatment (Paroxetine)	-0.6350	0.1263	-5.0300	<0.0001***
Time	-0.0064	0.0102	-0.6300	0.5260
Treatment (Paroxetine)×Time	0.0384	0.0097	3.9500	<0.0001***
Male Reproduction	GLMM with Poisson distribution and logarithmic link (groups = 5; observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.0042	0.0650		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 8)	2.1649	0.1171	18.4800	<0.0001***
Treatment (Paroxetine)	-0.5070	0.1866	-2.7170	0.0066**
Time	0.0189	0.0087	2.1620	0.0306*
Treatment (Paroxetine)×Time	-0.0074	0.0146	-0.5060	0.6129
Male Competitive Ability	GLMM with binomial distribution and logit link (groups = 5, observations = 122)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.0168	0.1295		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Intercept (Week 3)	-0.1096	0.1908	-0.5750	0.5660
Treatment (Paroxetine)	-1.1719	0.2850	-4.1130	<0.0001***
Time	-0.0017	0.0124	-0.1410	0.8880
Treatment (Paroxetine)×Time	0.0034	0.0194	0.1760	0.8600

* Indicates a p value <0.05, ** <0.01, *** <0.001.

3.10 Statistical Analyses

The time taken to produce first litter was analyzed with a Cox proportional hazard model (PH). As the gestation period is 21 days, any litters produced on or before day 41 were considered to be first litters. All breeding pairs that did not produce a litter at this time point were censored in the analysis. There were a total of 36 censorings and 24 events.

Litter size data are discrete counts and therefore were analyzed with a generalized linear mixed model (GLMM) with a Poisson distribution and logarithmic link. The model assessed the effects of treatment, parity, and the interaction of parity and treatment. The model intercept was set to litter one. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. There were 17 paroxetine-exposed litters from eight cages and 53 control litters from 20 cages used in this analysis.

Wean weight was analyzed with a linear mixed effects model (LMM). The model assessed the effects of treatment, parity, and the interaction of parity and treatment. The model intercept was set to litter one. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. Sexes were analyzed separately. A normal distribution was assumed, as weight data are continuous. Sample sizes used to assess weight differences include: 45 daughters from eight paroxetine-exposed breeding pairs compared to 89 daughters from 16 control breeding pairs and 23 sons from eight paroxetine-exposed breeding pairs compared to 88 sons from 17 control breeding pairs.

Body weight was analyzed with a LMM. The model assessed the effects of treatment, time, and the interaction of time and treatment on the 116 population founders. Sexes were analyzed separately. A normal distribution was assumed, as weight data are continuous. Treatment, time, and their interaction were modeled as fixed effects and individual and population were modeled as random effects with random slopes and intercepts generated for each. The intercept was set at week zero, as this was when founders were released into the enclosures. Collection of founder weight data continued on surviving individuals at each pup sweep. There were a total of 386 female observations from 76 mice and 179 male observations from 40 mice collected throughout the experiment.

Reproductive outputs were in terms of total offspring and thus are discrete data; therefore, these data were analyzed with a GLMM with a Poisson distribution and logarithmic link. The model assessed the effects of treatment, time, and the interaction of treatment and time on population-level reproduction. These effects were set as fixed effects in the model and population was set as a random effect with random intercept calculated for each. The intercept was set at week eight, as that was when the first collection period or pup sweep occurred. Reproductive output for each treatment was measured five times over the course of the 28-week study in each of the five independent populations for a total of 50 observations. Female reproductive output was analyzed in terms of total offspring and male reproductive success was analyzed in terms of male offspring. A post-hoc two-tailed Mann-Whitney U test was used to assess differences at each time point for female reproductive success as treatment was found to significantly, and oppositely, affect the intercept and slope of the model.

To assess the probability of territorial ownership, a GLMM was used. As a territory can be defended or not, a binomial distribution was used with a logit link. There were six territories within an enclosure and were either occupied by paroxetine-exposed males, control males or unoccupied. A total of 122 observations were collected from five populations and analyzed throughout the study. The model assessed the effect of treatment, time, and their interaction. These effects were set as fixed effects and population was set as a random effect with a random intercept generated for each. The model intercept was set to week three, as that was the first time point in which data existed in all five populations.

A multivariate PH was used to assess survivorship of founding individuals. Effects of treatment, population, and their interaction were examined in the model. Individuals that survived the length of the study or that were intentionally removed from the study were censored. Survival in females was assessed in 76 individuals: 14 events and 62 censorings. Survival in males was assessed in 40 individuals: 12 events and 28 censorings.

Proportional hazard models were conducted in JMP 9.0.3 (SAS institute Inc., Cary, NC) and two-tailed Mann-Whitney U tests were performed in Prism 5.03 (GraphPad Software Inc., La Jolla, CA). All mixed models were conducted in R 3.0.2 using either *glmer* or *lmer* functions within the lme4 library (Bates et al. 2014, R Development Core Team 2013). P values were calculated for LMM with the Swatterthwaite approximation under the lmerTest function (Kuznetsova et al. 2013). For all mixed models, several candidate models were fit to the data. These models varied in terms of random effects that estimated both intercept and/or slope. For each analysis, the

model that explained some of the variance with random effects and had the lowest Akaike Information Criterion (AIC) score was reported. The significance of treatment or the magnitude of the significance varied between models.

3.11 Supplementary Information References

Bates, D., M. Maechler, B. Bolker, S. Walker, R. Christensen, and H. Singmann. 2014. lme4: linear mixed-effects models using Eigen and S4. R package version:1.1-5.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Kuznetsova, A., P. Brockhoff, R. Christensen. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-3.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

CHAPTER 4

FACULTATIVE SEX RATIO ADJUSTMENT INCREASES FITNESS IN WILD MICE (*MUS MUSCULUS*) LIVING UNDER SEMINATURAL CONDITIONS

4.1 Abstract

Sex allocation theory suggests that parents who facultatively adjust their sex ratio towards the sex that is more likely to have high fitness (Trivers-Willard hypothesis) and/or towards the sex that is less energetically demanding to produce (cost of reproduction hypothesis) can increase their lifetime reproductive success. We investigated the expectations of both the Trivers-Willard hypothesis and the cost of reproduction hypothesis in mice whose sex ratios were biased towards females by exposure to the pharmaceutical paroxetine. Male offspring born to control dams weighed significantly greater than female offspring; however, no differences in wean weights were detected between male and female offspring from paroxetine-exposed dams. These data suggested that mothers did not save energy by skewing their litters towards females and does not support the cost of reproduction hypothesis. Offspring of paroxetine-exposed and control dams were released into seminatural enclosure where fitness of all individuals was determined. We found that female offspring from treatment dams had significantly higher reproductive success compared to the male offspring from the same litters relative

to controls. These data suggest that paroxetine-exposed dams were successful at increasing their lifetime reproductive success by facultatively adjusting litters towards female offspring and it is the first experimental evidence of determining the reproductive success of progeny from biased litters that supports the Trivers-Willard hypothesis.

4.2 Introduction

Perhaps one of the most influential hypotheses in sex ratio theory is the Trivers-Willard hypothesis (TWH). In polygynous species, a male of high quality can have substantial increases in reproductive success by mating with several females while low quality males may not breed at all. In contrast, reproductive success differences are not nearly as extreme in females, where both low and high-quality females typically produce similar numbers of offspring. Trivers and Willard (1973) suggest that high-quality mothers produce high-quality sons, who will out reproduce their sisters of the same condition; low-quality mothers produce low-quality daughters, who will out reproduce their brothers of the same condition. If parents are capable of facultative sex ratio adjustment of their offspring in response to environmental conditions, then parents will increase their lifetime reproductive success by increasing the rate at which their genes are passed onto future generations (Trivers and Willard 1973).

If differential investment occurs between female and male offspring, then parents can skew their litters in favor of the less expensive sex to reduce the cost of reproduction and increase their own rates of survival and lifetime fitness, especially when resources are limited; this idea is known as the cost of reproduction hypothesis (CRH) (Cockburn et al. 2002). The idea of CRH is similar to that of the TWH in that sex ratio adjustment can

lessen the risk of reproductive failure and will act to increase lifetime fitness. In terms of CRH, low quality mothers are reluctant to invest resources into the sex with the highest energy demands (often males in mammals). By preferentially investing into female offspring, the mother minimizes her own costs of reproductive efforts. When studying sex allocation it can often be difficult to disentangle these hypotheses as often times, there is supporting evidence for both hypotheses (Cockburn et al. 2002).

Experimental manipulations of diets to mimic poor ecological conditions have been successful in causing a bias in sex ratio and/or inducing differential investment into one offspring sex over the other in several rodent species. For example, when hamster dams were calorie restricted during pregnancy and lactation, litters were skewed towards female offspring and females weighed more than the male offspring (Labov et al. 1986). Likewise, when woodrat dams were calorie restricted during lactation, dams invested more into their female offspring than males as measured by offspring weight (McClure 1981). Mice consuming a low-calorie diet (Meikle and Drickamer 1986) or low-fat diet (Rivers and Crawford 1974, Rosenfeld and Roberts 2004) led to female biased litters. Although these studies provide evidence for CRH and TWH, these studies do not examine the fitness of the offspring from these biased litters to determine if mothers gain reproductive benefits by skewing the sex ratio.

The classic ecological example of TWH is the case of the red deer (*Cervus elaphus*), where dominant females are more likely to produce male offspring than subordinate females, as sons produced by dominant females have higher fitness than do their daughters (Clutton-Brock et al. 1984, 1986). Interestingly, more recent work on this population has revealed that as population densities increase, the biasing of sex ratios

towards sons in dominant females has disappeared due to the increased difficulty in raising offspring under these conditions. This result, paired with the observation that male offspring are more likely to die under low resource conditions (i.e., they are more energetically expensive to successfully rear), indicates that the CRH is also at play in explaining the sex ratios of this population (Kruuk et al. 1999). In fact, due to the countervailing CRH and TWH pressures on sex ratio it has been argued that the original observation of bias would not have been made if the original study started in the1990s (Cockburn et al. 2002).

In a previous experiment, we assessed the safety of the antidepressant paroxetine [Paxil®, GlaxoSmithKline (GSK), Brentford, England] in wild-derived house mice (*Mus musculus*). We exposed mice to ~1.82 mg/kg/d of paroxetine starting *in utero* and ending in early adulthood. These animals were removed from the treatment and then released into seminatural enclosures where they competed against control animals for mates, resources, and territories over a 28-week period. We reported that within these enclosures, paroxetine-exposed males weighed 13% less, dominated 53% fewer territories, and had 44% less offspring relative to controls; likewise, paroxetine-exposed females had an initial decline in reproduction relative to controls but rebounded at later time points (Chapter 3). We refer to this seminatural mouse model system as organismal performance assays (OPAs) and have used to quantify fitness costs of several genetic, nutritional, and pharmacological treatments (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013, Ruff et al. *in review*, Chapter 2, Chapter 3, Chapter 5).

Here we analyze the data from the above study in the context of both the TWH and CRH. All comparisons within this paper are unique from those occurring in Chapter 3. First, we assess whether paroxetine-exposure (from here on referred to as treatment) causes biases in the sex ratio of litters relative to theoretical values and control litters. Second, we determine if treatment affects litter size. Third, we test if there is differential investment in female and male offspring, measured by offspring wean mass, and if the degree of this differential investment differs between treatment and control dams. Fourth, we determine if female and male offspring of treatment dams have differential lifetime fitness (reproductive success) in seminatural enclosures relative to controls. Finally, we aim to calculate the fitness benefits achieved through facultative sex ratio adjustment when sex ratio skew is detected.

4.3 Materials and Methods

4.3.1 Animals. Wild-derived genetically outbred house mice were used in this experiment as they exhibit natural behavioral characteristics that allow them to function in natural and seminatural environments, unlike inbreed mouse strains (Nelson et al. 2013). In this experiment, animals were from the 12th generation of the colony that is described by (Meagher et al. 2000). Genetic diversity was assessed in the 11th generation and found to be comparable to that in the wild (Cunningham et al. 2013). Individuals were provided access to food and water *ad libitum* and maintained on a 12:12 hr light:dark cycle. All procedures were approved by the University of Utah IACUC.

4.3.2 Treatment. Paroxetine (GlaxoSmithKline, US, Molecular formula: $C_{19}H_{20}FNO_3 \cdot HCl$) was administered via diet at a dose of ~1.82 mg/kg/d and has been

extensively described in Chapter 3. Sixty breeder pairs were selected for this experiment, 20 treatment and 40 controls. The asymmetry in cage numbers is because control animals were used for both this and another experiment. When paired, female breeders were 75.8 ± 13.2 (M ± SD) weeks old and male breeders were 70.4 ± 13.8 weeks old. All animals were individually housed and females were fed their respective diets eight days prior and males five days prior to breeding. Breeders were remained together until a maximum of four litters were produced. At 28 days of age, litters were weaned and housed with same sex siblings. Upon weaning, we recorded individual mass, litter size and sex. After weaning, treatment continued until animals were released into the seminatural enclosures, at which all animals were provided with the control diet *ad libitum*.

4.3.3 Seminatural enclosures. Enclosures have previously been described in (Ruff et al. 2013). Briefly, enclosures are ~30 m^2 and contained four defendable optimal territories with multiple dark nesting sites and direct access to food. In addition, each enclosure consisted of two suboptimal territories with nesting boxes exposed to light and had indirect access to food. Territories were separated by hardware mesh that was easily climbed but added an element of spatial complexity. Food and water were provided *ad libitum* in all territories. Animals were kept on a 12:12 hr light:dark cycle.

Five independent populations were established and maintained for 28 weeks. Populations consisted of eight males and between 14-16 females; these animals are referred to as founders for a total of 116 animals. Half of each sex were treatment individuals and the remainders were controls. This population structure allows treatment individuals to directly compete with controls for resources, territories, and mates.

Enclosure space and population size created a population density reported within the range observed in the wild (Sage 1981).

Upon enclosure release, male mice were on average 14.3 ± 4.7 (M ± SD) weeks old and females were 18.9 ± 6.9 weeks old. Males were released into enclosure with non-experimental females to allow males to establish territories and prevent incidental breeding with experimental females. Nonexperimental females were removed and replaced with experimental females one week later. Although unrelatedness at the cousin level or above was ideal for founders, four populations consisted of one to two brother pairs and all populations consisted of sister pairs and four populations consisted of sister triplets. At most, one male, per treatment or control group, was in a population containing two of his sisters. When relatedness was present, it was balanced between treatment and control groups.

4.3.4 Reproductive success within seminatural enclosures. Reproductive success of founders was determined by removing and genotyping all offspring during five-week intervals within seminatural enclosures and have been described in Chapter 3. Briefly, reproductive success was determined on a population-level by examining sex-specific allelic variants (Y-chromosome and mitochondrial genome) in three of the five populations while the remaining two populations were assessed on an individual level by examining multiple autosomal microsatellite loci. Data generated from the multiple satellite analysis were converted to population-level readouts (total number of offspring per treatment or control groups within a population versus total number of offspring per founder). A total of 872 samples were collected with an average of 174.4 ± 38.1 (M ± SD) offspring per population. Mitochondrial genotypes were assessed in 626 samples

(three of five populations) and obtained for 100% of offspring. Y-chromosome genotypes were assessed in all five populations to determine reproductive output and the sex ratio of offspring. Of the 872 offspring, 414 Y- chromosome genotypes were obtained, suggesting that 95% of all males were typed assuming the sex ratio was 1:1. For the two populations that underwent parentage analysis, mother-father-offspring relationships were accepted when 95% trio confidence was obtained in Cervus 3.0 (Kalinowski et al. 2007). Using this rule, 91% (187/205) of one population was genotyped and 75% (147/195) of the second population was genotyped.

4.4 Statistical Analyses

As offspring can only be female or male, sex ratio was analyzed with a generalized linear mixed model (GLMM) with a binomial distribution and logit link as is recommended (Wilson and Hardy 2002). Birth cage was set as a random effect to control for repeated measures as multiple litters (up to four for each breeding pair) were assessed. Several candidate models were generated including or excluding litter parity and its potential interaction; the final model was selected based the lowest upon AIC score. There were 70 observations (litters) from 28 groups (cages).

As reproduction data are discrete counts, we modeled offspring counts across treatment and control breeding cages in a GLMM with a Poisson distribution and a logarithmic link. The model assessed the main effects of treatment and litter parity and their interaction on litter size. There were 70 observations (litters) made in 28 groups (breeding cages). Treatment, litter parity, and their interaction were modeled as fixed effects and breeding cage was modeled as a random effect. Several candidate models

were generated, including or excluding parity and its potential interaction; the final model was selected upon the lowest AIC score.

For mass comparisons, a linear mixed-effects model (LMM) was used to assess the main effects of treatment, sex, and litter parity, as well as their respective interactions on the weaning mass of 245 observations (pups) from 25 groups (breeding cages). Treatment, sex, litter parity, and their interactions were modeled as fixed effects and breeding cage was modeled as a random effect to control for repeated measures as multiple litters (up to four) were assessed. Several candidate models were generated treating including or excluding parity and its potential interactions; the final model was selected based upon the lowest AIC score.

As offspring can only be female or male, sex ratio was analyzed with a GLMM with a binomial distribution and logit link. Reproductive output for each treatment was measured five times over the course of the 28-week study in each of the five independent populations for a total of 50 observations. Population was set as a random effect to control for repeated measures and treatment, time, and their interaction were modeled as fixed effects. Several candidate models were generated, including or excluding time. The final model was selected based on the lowest AIC score.

A GLMM with a binomial distribution and logit link was used to determine the likelihood of a given offspring born within seminatural enclosures as being produced by a female or male offspring from a treatment dam relative to being produced by a female or male offspring from a control dam. This method of analysis was selected as raw reproduction in enclosures is based on separate measurement for each sex—mitochondrial and multiple autosomal microsatellite loci for females and Y-chromosome

marker for males. By comparing the relative success of treatment female and males relative to their control counterparts, comparisons between the relative fitness of treatment females and males can be made. Sex, time, and their interaction were modeled as fixed effects and population was modeled as a random effect to account for repeated measures. Reproductive output was measured five times over the course of the 28-week study in each of the five independent populations for a total of 50 observations. Several candidate models were generated, including or excluding time, and the final model was selected based on the lowest AIC score.

All GLMMs and LMMs were conducted in R 3.0.2 using the lme4 library (R Development Core Team 2013, Bates et al. 2014). Degrees of freedom and resulting p-values for LMMs were determined with a Satterthwaite approximation using the lmerTest library (Kuznetsova et al. 2013). As controversy exists in how to best calculate degrees of freedom in LMMs, it should be noted that all effects deemed statistically significant through their resulting p-values also possess a t-value > |2|, a conservative criteria for significance recommended by the library's authors. All alpha values were 0.05 and tests were two-tailed.

4.5 Results

Sex ratio was affected by treatment as paroxetine-exposed breeders produced litters with a sex ratio of 32:68 M:F, while control breeders had a sex ratio of 51:49 M:F. Sex ratio of treatment breeders was skewed towards female offspring when compared with an expected 50:50 ratio (GLMM; $z = -2.76$, $p < 0.01$; Figure 4.1A) and when

compared to control breeders (GLMM; $z = 2.39$, $p < 0.05$). For a complete readout of mixed model results for breeding cage sex ratio, litter size, and wean mass, see Table 4.1.

Although significant bias towards male offspring was detected in the treatment group, no differences in litter size were observed between treatment and control breeders (GLMM; $z = 0.18$, $p = 0.86$; Figure 4.1B). Treatment breeders had an average of 4.15 (SEM + 0.62, −0.53) pups per litter while control breeders had an average of 4.28 (+ 0.46, −0.63) pups per litter. SEMs are asymmetric, as values have been back-transformed from logarithmic data.

Treatment offspring had reduced mass at weaning than their control counterparts (LMM; $t = -3.40$, $p < 0.001$; Figure 4.2). In terms of wean mass, control dams invested more into their male offspring than their female offspring as male offspring from control dams had a mass that was 12.5% (1.54 ± 0.19 g) greater than female offspring at weaning (LMM; $t = 7.98$, $p < 0.0001$). Differential investment between male and female offspring by treatment dams was reduced relative to controls (LMM; $t = -2.68$, $p < 0.01$) as male offspring only had a mass that was 4% (0.52 ± 0.38 g) greater that female offspring. Litter parity affected wean mass where individuals from subsequent litters were on average 0.50 ± 0.16 g heavier (LMM; $t = 3.23$, $p < 0.01$). As a significant effect of treatment, sex, and their interaction was detected in the LMM, a post-hoc Tukey's test was performed on all pair-wise combinations and all were found to be significant ($p < 0.05$), except between the masses of female and male offspring from treatment cages ($p = 0.82$), indicating that treatment dams invest equally in regards to the sex of their offspring.

In seminatural enclosures, treatment and control founders had an overall sex ratio of 44:56 M:F. Sex ratio of treatment founders was skewed towards female offspring (GLMM; $z = -2.53$, $p < 0.05$) when compared with an expected 50:50 ratio. However, they did not differ from those of control founders (GLMM; $z = 0.19$, $p = 0.85$), suggesting that the overall sex ratio within the seminatural enclosures was biased towards females, but did not differ between groups. For a complete readout of GLMM results for sex ratio and overall reproductive success comparisons in seminatural enclosures, see Table 4.2.

Female and male treatment founders had reduced reproductive success when compared with their control counterparts. The probability that a pup was born to a treatment female founder was significantly higher than that of a pup that was sired by a treatment male founder (GLMM; $z = -2.98$, $p < 0.001$; Figure 4.3), relative to control founders. These data suggest that treatment male founders suffered reduced reproductive success when compared to their female counterparts, relative to control founders.

4.6 Discussion

Treatment with paroxetine caused substantial sex ratio skewing relative to both symmetric theoretical expectations and control breeding cages. Sex ratios within treatment litters were ~1:2 M:F and only one litter out of 17 had a sex ratio of greater than 50% male. Sex-ratio biases were not detected in other rodent studies assessing the safety and bioactivity of paroxetine possibly due to the use of laboratory mouse strains (Coleman et al. 1999), which have undergone generations of selection for consistent and reliable reproduction in cages. As this study used wild-derived mice that have been bred

to maintain genetic diversity, it is possible that they still possess the, as of yet identified, molecular machinery necessary to skew the sex ratio of litters based on the stimulus of an external factor such as paroxetine. It has yet to be determined if paroxetine exposure consistently skews sex ratios of mice across factors such as age, dose, length of exposure, and maternal body weight, but importantly, in the case outlined in this study, it resulted in dramatic sex ratio biases. Interestingly, no differences in sex ratios were detected between treatment and control founders in seminatural enclosures though both groups had significantly fewer males than expected values. The female bias detected in OPAs may be due to the stressful seminatural environment but is more likely an artifact as our genotyping system is biased to undercounting male offspring.

Sex ratio was significantly affected by the exposure of paroxetine while no differences in litter size was detected when compared to control litters. Although mechanisms were not evaluated in this study, Rosenfeld and Roberts (2004) propose four nonmutually exclusive theoretical hypotheses to explain mechanisms of sex ratio skewing in mammals. Prior to conception, two hypotheses for sex ratio bias are 1) differential sperm motility of one class of sperm over the other and 2) one class of sperm may be more effective at fertilizing the egg over the other. These two hypotheses are suggested to be female mediated as conditions in the reproductive tract can drive this selection. It is possible that paroxetine exposures altered the conditions of the female reproductive tract that favored X-bearing sperm over Y-bearing sperm. The remaining two hypotheses pertain to sex ratio biasing after fertilization has occurred. Hypothesis 3 suggests a difference in the rate at which XX and XY embryos develop prior to implantation. In litter bearing mammals, faster developing embryos have a competitive advantage when it

comes to securing a place of attachment to the uterine wall. As male mice have a higher development rate (Valdivia et al. 1993), this hypothesis is unlikely to be supported in our study. Hypothesis 4 suggests an increased arrested development in one embryo sex over the other. Perhaps males were selectively aborted because they have a faster development rate and paroxetine–exposed dams were incapable of sustaining the energetic demands of producing male offspring; however, this selective abortion may lead to a reduced litter size, which was not detected in our study. Though our findings offer support for some of these hypotheses concerning the mechanistic cause of sex ratio adjustment, further studies are needed to elucidate the mechanisms of paroxetine-mediated sex ratio bias.

Though male mice typically have greater masses than their female littermates (Chapter 2, Chapter 5), this was not the case for offspring within treatment litters. This finding indicates that the CRH does not explain the observed sex ratio biasing, as the overproduction of female offspring did not result in an energetic savings for treatment dams. The decreased mass of treatment offspring in general relative to controls is an indication that treatment dams were in less than optimal condition and unable to invest maximally in their litters and therefore, the observed weight data support the first requirement of the TWH by indicating that treatment dams are of low quality. As a caveat, it should be mentioned that wean mass is a rather gross measure of maternal investment, especially as offspring typically begin consuming solid food around day 17, although nursing has been documented up until day 28 (Konig and Markl 1987), which is when these animals were weaned. However, so long as there is a correlation between maternal investment and wean mass, then these data should at least, in part, address the question of maternal investment as a function of sex ratio.

In seminatural enclosures, treatment founders competed directly with control founders and though treatment founders of both sexes performed more poorly than their control counterparts this difference was more pronounced in males, lending support to the TWH as an explanation for the observed sex ratio biasing. Female treatment founders only had 19% fewer offspring while male treatment founders had 44% fewer offspring than control founders, which lead to a situation where offspring produced in seminatural enclosures were much more likely to be born to female treatment founders than to male treatment founders. The differential fitness between male and female treatment founders is likely due to the high levels of male intrasexual competition within polygynous mating systems, found within house mice societies. Within semi-enclosures, it has been demonstrated that territorial dominant males sire ~80% of offspring within an enclosure and that treatments that decrease relative rates of territorial acquisition also decrease fitness (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013, Chapter 2, Chapter 3). As male treatment founders acquired only half as many territories as controls, it is likely that this discrepancy in competitive ability explains the fitness differences observed between treatment and control male groups and underscores the relative fitness of treatment-born females *versus* males (Chapter 3). These observations support the TWH as an explanation for the observed sex ratio bias as low-quality (treatment) mothers produced low-quality sons, who were out reproduced by their sisters of the same condition.

Taken together, the presence of substantial skew in sex ratio towards female offspring in treatment breeding cages, and the observation that these females had higher fitness than males from the same litters, provides an example of adaptive facultative sex

ratio bias in mice. By measuring the fitness of both female and male treatment founders relative to controls in seminatural enclosures, we can envision that if a treatment dam only had female offspring, then her fitness would be ~81% that of control female and that if she only had male offspring, her relative fitness would be ~56%. With these extreme cases described, a prediction can be made estimating the second-generation fitness of treatment dams relative to controls across varying litter sex ratios (Figure 4.5). It can be extrapolated that at the ~1:2 M:F sex ratio observed in treatment litters, dams increased their second generational fitness by ~ 5.4% than they would have had with litters consisting of a 1:1 sex ratio. Litters containing 0% males would have led to an increase in second-generational fitness to 81% and these dams would have reclaimed the maximum possible amount of fitness (100%). Litters containing 100% male offspring would have lead to a decreased in second-generational fitness to 56% and these dams would have lost the most fitness (−100%). Treatment breeders were successful at reclaiming ~ 47% of the maximum fitness that could have been obtained by skewing their ratios away from 1:1 towards ~1:2 M:F.

Here we have demonstrated that a pharmaceutical agent can dramatically bias sex ratios in wild-derived mice and that this facultative biasing is adaptive. Furthermore, by assessing maternal investment through wean mass measures and offspring fitness, we have determined that TWH and not the CRH best explains our observations as biasing does not result in energetic savings to dams, but does result in increased second-generational fitness. This is also the first experimental evidence determining second-generational fitness that supports TWH. Though facultative sex ratio biasing is common across animal taxa, experimental demonstrations within mammals have proven elusive;

our seminatural model system provides an ecologically relevant fitness assay for a well-studied mammal that could be applied across a battery of treatments documented to induce sex ratio biasing to determine their adaptive nature. Furthermore, the demonstration that paroxetine exposure induces sex ratio biasing towards female mice provides a new system in which to pursue the enigmatic mechanism, which underlay the process.

4.7 Acknowledgments

I thank my PhD advisor, Wayne Potts for his support; James Ruff for his significant contribution to this chapter; Michelle Bartlett, Stephanie Eddy, Tessa Galland, Kristie Kandaris, David Kircher, Nicole Liu, Linda Morrison, Elyse Schwab, Mirtha Sosa, Tristan Underwood, and Elizabeth Young for enclosure maintenance, data collection, and genotyping; Ahmed Bwika and Jon Gale for animal care. This project was funded with support from the University of Utah's Technology Commercialization Program and was partially conducted while Wayne Potts was supported by NSF grant DEB 09-18969 and NIH grant R01-GM109500. I was supported by an NSF GK-12 Educational Outreach Fellowship (DGE 08-41233).

4.8 References

Bates, D., M. Maechler, B. Bolker, S. Walker, R. Christensen, and H. Singmann. 2014. lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-5.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness. 1984. Maternal dominance, breeding success and birth sex ratios in red deer. Nature 308:358–60.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness. 1986. Great expectations: dominance, breeding success and offspring sex ratio in red deer. Animal Behaviour 34:460–471.

Cockburn, A., S. Legge, and M. Double. 2002. Sex ratios in birds and mammals: can the hypotheses be disentangled? *in* Sex Ratios Concepts and Research Methods. Pages 266–286 (I. Hardy, Ed.). Cambridge University Press, Cambridge, UK.

Coleman, F. H., H. D. Christensen, C. L. Gonzalez, and W. F. Rayburn. 1999. Behavioral changes in developing mice after prenatal exposure to paroxetine (Paxil). American journal of obstetrics and gynecology 181:1166–71.

Cunningham, C. B., J. S. Ruff, K. Chase, W. K. Potts, and D. R. Carrier. 2013. Competitive ability in male house mice (Mus musculus): genetic influences. Behavior genetics 43:151–60.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular ecology 16:1099–106.

Konig, B., and H. Markl. 1987. Maternal care in house mice. Behavioral Ecology and Sociobiology 20:1–9.

Kruuk, L. E., T. H. Clutton-Brock, S. D. Albon, J. M. Pemberton, and F. E. Guinness. 1999. Population density affects sex ratio variation in red deer. Nature 399:459–61.

Kuznetsova, A., R. Brockhoff, and R. Christensen. 2013. lmerTest: Tests for random and fixed effects for linear mixed effects models(lmer objects of lme4 package). R package version 2.0-3.

Labov, J. B., U. W. Huck, P. Vaswani, and R. D. Lisk. 1986. Sex ratio manipulation and decreased growth of male offspring of undernourished golden hamsters (Mesocricetus auratus). Behavioral Ecology and Sociobiology 18:241–9.

McClure, P. A. 1981. Sex-Biased Litter Reduction in Food-Restricted Wood Rats (Neotoma floridana). Science 211:1058–60.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

Meikle, D. B., and L. C. Drickamer. 1986. Food availability and secondary sex ratio variation in wild and laboratory house mice (Mus musculus). Journal of Reroduction and Fertility 78:587–91.

Nelson, A. C., J. W. Cauceglia, S. D. Merkley, N. A. Youngson, A. J. Oler, R. J. Nelson, B. R. Cairns, E. Whitelaw, and W. K. Potts. 2013. Reintroducing domesticated wild mice to sociality induces adaptive transgenerational effects on MUP expression. Proceedings of the National Academy of Sciences 110:19848–53.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rivers, J. P., and M. A. Crawford. 1974. Maternal nutrition and the sex ratio at birth. Nature 252:297–8.

Rosenfeld, C. S., and R. M. Roberts. 2004. Maternal diet and other factors affecting offspring sex ratio: a review. Biology of reproduction 71:1063–70.

Ruff, J., S. Hugentobler, A. Suchy, M. Sosa, R. Tanner, L. Morrison, S. Gieng, M. Shigenaga, and W. Potts. (in review). Sugars within high fructose corn syrup reduce survival and fitness of female mice compared to surcrose. The Journal of the federation of American societies for experimental biology.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

Sage, R. 1981. The Mouse in Biomedical Research. Pages 40–90 (H. Foster, J. Small, and J. Fox, Eds.). Academic Press, New York.

Trivers, R. L., and D. E. Willard. 1973. Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring. Science 179:90–2.

Valdivia, R. P., T. Kunieda, S. Azuma, and Y. Toyoda. 1993. PCR sexing and developmental rate differences in preimplantation mouse embryos fertilized and cultured in vitro. Molecular reproduction and development 35:121–6.

Wilson, K., and I. C. W. Hardy. 2002. Statistical analysis of sex ratios: an introduction. *in* Sex Ratios Concepts and Research Methods. Pages 48–92 (I., Hardy, Ed.). Cambridge University Press, Cambridge, UK.

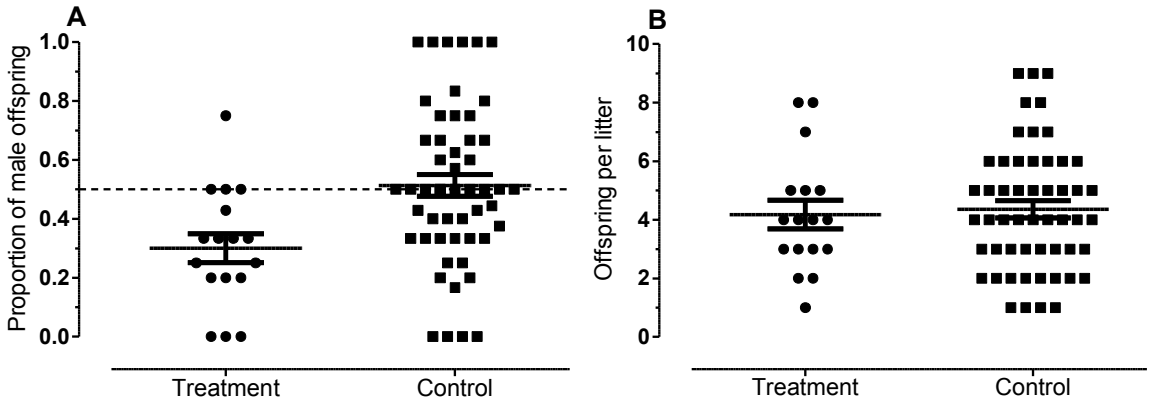

Figure 4.1. Sex ratio of litters and litter size from treatment and control breeding cages. A) Litters from treatment breeders were female biased and significantly deviated from the expected 1:1 sex ratio, [*n* = 28 cages, *observations* = 70 litters (GLMM; z = −2.78, p < 0.01)]. Likewise, the sex ratio of treatment litters differed significantly from that of controls [*n* = 28 cages, *observations* = 70 litters (GLMM, z = 2.39, p < 0.05)]. **B)** No differences in litter size were detected between treatment and control breeders [*n* = 28 cages, *observations* = 70 litters (GLMM; z = 0.18, p = 0.86)]. Lines represent means and error bars, standard error.

Table 4.1. Mixed model results for breeding cage litter size, sex ratio, and wean mass.

Sex ratio of litters born in cages	GLMM with binomial distribution and logit link (groups = 28, observations = 70)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.0463	0.2151		
Fixed Effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Treatment (Intercept)	-0.7426	0.2670	-2.7810	0.0054**
Control	0.7239	0.3028	2.3910	0.0168*
Litter size	GLMM with binomial distribution and logit link (groups = 28, observations = 70)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.0344	0.1856		
Fixed Effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Treatment (Intercept)	1.4247	0.1380	10.3240	<0.0001***
Control	0.0284	0.1597	0.1780	0.8590
Wean mass of treatment and control offspring by sex	LMM (groups = 25, observations = 245)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	2.0225	1.4221		
Cage (Slope)	0.2195	0.4686		
Litter Size (Intercept)	0.3360	0.5797		
Fixed Effects	*Estimate*	*Std. Error*	*t value*	*Pr(> | t |)*
Intercept (Control)	10.0390	0.4799	20.9180	<0.0001***
Treatment	-1.5356	0.4521	-3.3960	0.0023**
Sex (Male)	1.5376	0.1928	7.9770	<0.0001***
Sex (Male)*Treatment	1.0165	0.3795	-2.6780	0.0080**
Parity	0.5026	0.1558	3.2270	0.0066**

* Indicates a p value < 0.05, ** < 0.01, *** < 0.0001.

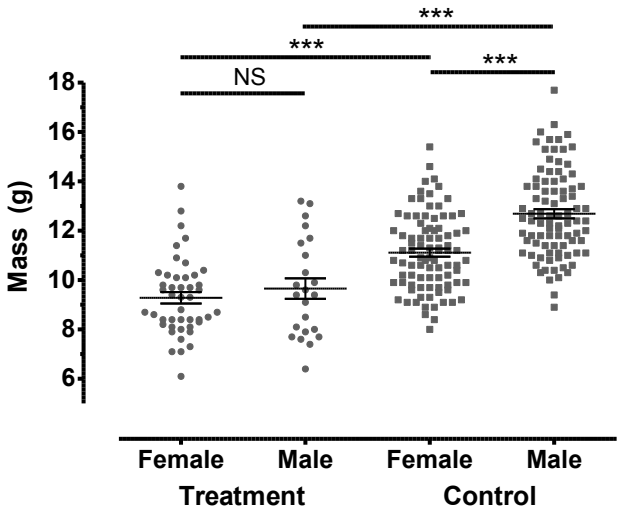

Figure 4.2 Maternal investment in treatment and control offspring as measured by wean mass. Treatment pups of both sexes had reduced wean mass when compared with their control counterparts [*n* = 25 cages, *observations* = 245 offspring (Tukey's post-hoc test; female, $p < 0.001$; male, $p < 0.001$)]. In control cages, male offspring had a mass of 1.54 g greater than female offspring ($p < 0.001$). However, in treatment cages, male and female offspring did not have different masses ($p = 0.82$). Lines represent means and error bars, standard error of raw data and *** indicate $p < 0.001$.

Table 4.2. Generalized linear mixed model results for sex ratio and overall reproductive success comparisons within seminatural enclosures.

Sex ratio of offspring from seminatural enclosures	GLMM with binomial distribution and logit link (groups = 5, observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.0000	0.0000		
Fixed Effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Treatment (Intercept)	-0.2592	0.1024	-2.5310	0.0114*
Control	0.0267	0.1373	0.1950	0.8457
Reproductive output comparisons of female and male treatment founders relative to controls	GLMM with binomial distribution and logit link (groups = 5, observations = 50)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.1838	0.4287		
Fixed Effects	*Estimate*	*Std. Error*	*z value*	*Pr(> | z |)*
Female Founders	-0.2109	0.2042	-1.0330	0.3017
Male Founders	-0.3742	0.1256	-2.9800	0.0029**

* Indicates a p value < 0.05, ** < 0.01.

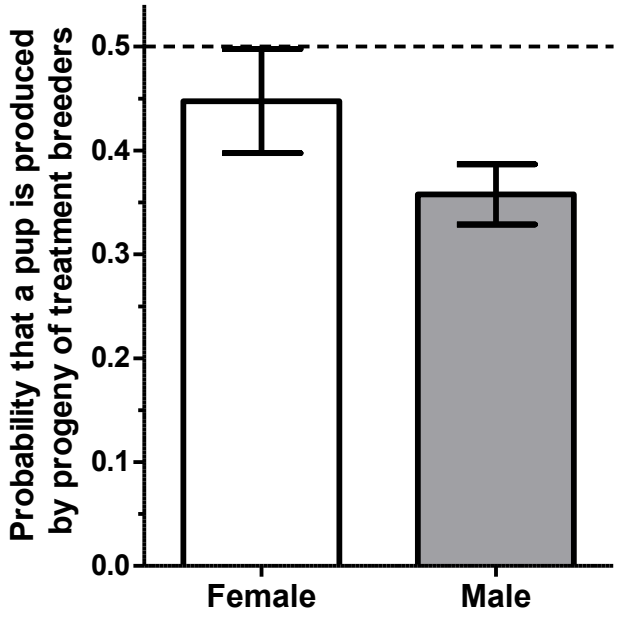

Figure 4.3. Fitness of female and male treatment founders relative to controls. The probability that a pup was born to a female treatment founder was significantly higher than the probability of a pup being sired by a male treatment founder [n = 5 populations, *observations* = 50 (GLMM; $z = -2.98$, $p < 0.01$)]. The dashed line indicates the level of fitness required to be equivalent to controls. Error bars represent standard error from five independent populations in seminatural enclosures.

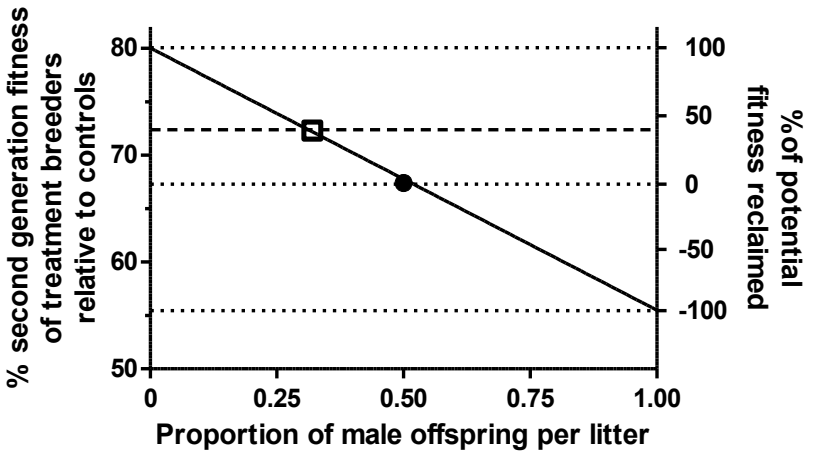

Figure 4.4. Influence of litter sex ratio on second-generation fitness of treatment breeders relative to controls (left y-axis) and the percentage of possible fitness recovered by facultative sex ratio adjustment (right y-axis). Within seminatural enclosures, total offspring of treatment breeders had 67% of the fitness of control breeders (dashed line); however, as the fitness of male offspring to female offspring from treatment cages significantly differed, treatment dams would have increased their relative fitness to achieve ~81% if they had produced litters with 0% males (top dotted line). On the contrary, treatment dams would have decreased their relative fitness to 56% if they produced litters with 100% males (bottom dotted line). These extremes are indicated by the top and bottom dotted lines and inform the 100% levels of the right y-axis. If treatment breeders did not skew their sex ratio and had litters consisted of a 1:1 sex ratio, then their second-generation fitness would have decreased by ~5.4%, depicted by the closed circle and middle dotted line. The empty square represents the expected fitness achieved by treatment breeders at the observed sex ratios. At the level of sex ratio bias detected treatment breeders, they were capable of reclaiming ~ 47% of their potential fitness gains by facultatively adjusting their litters towards female offspring.

CHAPTER 5

ORANISMAL PERFORMANCE ASSAY REVEALS INCREASED FEMALE REPRODUCTION WHEN EXPOSED TO ROFECOXIB

5.1 Abstract

We utilized a novel toxicity quantification method known as the organismal performance assay (OPA) to assess the safety of rofecoxib, a drug withdrawn from the market after being linked with causing more than 27,000 cardiac events. OPAs utilize genetically diverse wild mice (*Mus musculus*) that compete amongst each other for limited resources in seminatural enclosures. Performance measures included reproductive success, survivorship, and male competitive ability. Within OPAs, rofecoxib-exposed males did not suffer performance declines when compared with controls and had equal reproduction, competitive ability, and survivorship. Rofecoxib-exposed females experienced a 40% higher reproductive output compared to control females. The adverse health effects of rofecoxib seen in humans escaped the detection of OPAs and may have been because animals in OPAs were no longer on rofecoxib exposure and no cumulative damage was caused from previous exposure. This is the second study in which OPAs did not detect adversity and therefore, OPAs would still be useful if implemented during preclinical studies and other areas of toxicity assessment.

5.2 Introduction

Rofecoxib (Vioxx®, Merck; Whitehouse Station, NJ) is a selective nonsteroidal anti-inflammatory drug (NSAID) that was prescribed to patients to relieve arthritic pain. The FDA approved rofecoxib in 1999, but it was recalled by Merck in 2004 after the drug was shown to increase the risk of cardiovascular events when patients took the drug >18 months (Keane 2004). During its time on the market, ~107 million prescriptions of rofecoxib were dispensed (Krumholz et al. 2007), to 80 million patients (Jüni et al. 2004), generating a revenue of $5 billion dollars in sales (Drazen 2005). More than 27,000 cardiac events were associated with the use of rofecoxib (Horton 2004) and Merck has paid more than $6 billion in lawsuit settlements and legal fees (Loftus 2010).

Rofecoxib was targeted to reduced prostaglandin synthesis by selectively inhibiting cyclo-oxygenase-2 (COX-2) enzymes, which are widely expressed at sites of inflammation (Jüni et al. 2004). COX-1 enzymes are expressed in many tissues, including gastrointestinal tract (Jüni et al. 2004, Claycomb et al. 2011). Prolonged use of non-selective NSAIDs, such as naproxen and ibuprofen, can cause gastrointestinal toxicity because these drugs inhibit both COX-1 and COX-2 enzymes (Jüni et al. 2004). Prostaglandins act as a vasodilator and inhibit platelet aggregation while thromboxane (synthesized by COX-1) has the opposite effect, but both are important in maintaining cardiovascular homeostasis (Ricciotti and FitzGerald 2011). The imbalance of prostaglandins and thromboxane is the suggested mechanism of rofecoxib-induced cardiac adversity (Krumholz et al. 2007).

Preclinical studies were not geared at assessing cardiac safety of rofecoxib. Rather, preclinical studies determine if a pharmaceutical of interest causes mutagenicity,

carcinogenicity, teratogenicity, and infertility (Merck 2002). Rofecoxib exposure was not found mutagenic in rodent cells, nor was it found to be carcinogenic, teratogenic, or caused infertility in rodents (Merck 2002). Cardiac adversity was first suspected after market release during a phase IV clinical trial Vioxx gastrointestinal outcome research (VIGOR) study (Bombardier et al. 2000). These results were later confirmed by the adenomatous polyp prevention on Vioxx (APPROVe) study, which ultimately lead to the recall of rofecoxib (Keane 2004).

During human clinical trials, 73% of pharmaceutical that have passed preclinical safety assessments fail (Lipsky and Sharp 2001) and 10% of FDA approved pharmaceuticals are recalled after market release due to unforeseen toxicity (Schuster et al. 2005). The financial costs associated with pharmaceutical development are ~$1.4 billion per compound and require 12-15 years of research per drug (Miller 2012). Pharmaceuticals that fail after FDA approval can cause substantial costs associated with litigation fees, such as rofecoxib. One potential solution to reduce human suffering and to decrease the high failure rate of approved pharmaceuticals is to employ new methodologies in preclinical assessment that provide results without ambiguous interpretation.

We have developed a novel toxicity assay, known as the organismal performance assay (OPA), which has the potential to be valuable if implemented during preclinical testing. OPAs are comprised of wild mice (*Mus musculus*) from either a treatment group or a control group that compete amongst each other for resources in seminatural enclosures. Performance of individuals is measured on an ultimate level in terms of Darwinian fitness (i.e., life long reproduction) and its key components (e.g., social

dominance and survival). OPAs are capable of revealing mammalian toxicity with high sensitivity because this approach challenges most physiological systems synergistically and in order for individuals to be successful, high performance from most physiological systems is required. OPAs have previously revealed their sensitivity by quantifying the effects of cousin and sibling-level inbreeding, harboring a selfish gene, and consuming added sugar at human relevant levels (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013), wherein all of these studies, OPAs found substantial deleterious effects that were missed by current methodologies.

Here we use OPAs to determine if rofecoxib exposure causes fitness declines in wild mice. Our findings will address the current lack of information concerning rofecoxib exposure on ultimate measures (i.e., fitness) of organismal health. If rofecoxib exposure adversely affects any physiological system, we predict that exposed individuals will suffer survival, dominance, and reproduction declines relative to controls.

5.3 Materials and Methods

5.3.1 Animals. Genetically diverse, wild-derived house mice (*Mus musculus*) were used in this experiment. Wild mice possess natural behaviors needed to function within seminatural environments (Nelson et al. 2013). The wild mice used in this experiment were from the 12th generation of the colony described by (Meagher et al. 2000). Genetic diversity of this colony was assessed in the 11th generation and found to be comparable to wild populations (Cunningham et al. 2013). Animals were provided food and water *ad libitum* and maintained on a 12:12 hr light:dark cycle. All procedures and protocols were approved by the University of Utah IACUC.

5.3.2 Drug exposure. Dosing was achieved by incorporating 12.5 g of rofecoxib (AK Scientific Inc. Union City, CA; Molecular formula: $C_{17}H_{14}O_4S$) into 50 kg of rodent chow (TD.130006; Harlan Teklad, Madison, WI). As wild mice eat approximately 3 g per day and weigh 20 g (Reagan-Shaw et al. 2008), individuals will ingest 0.75 mg of rofecoxib per day or 37.5 mg/kg/day. Using a metabolic rate conversion factor, this is equivalent to a human dose of 3.0 mg/kg/day, or a daily dose of 182.4 mg, assuming the average human weighs 60 kg (Reagan-Shaw et al. 2008). Rofecoxib was prescribed at doses 12.5 – 50 mg/d (Merck 2002), thus animals in this experiment were exposed to a dose 3.5 fold higher than human therapeutic dose but it is within the range of doses in preclinical studies (Merck 2002).

Sixty breeding pairs were selected for this experiment and divided into two treatments: rofecoxib-exposed or control. All breeders were individual housed eight days prior to pairing. Females in the rofecoxib treatment started exposure at this time while males in the rofecoxib treatment started exposure five days prior to pairing. Breeding pairs were kept together until a maximum of three litters were produced. All offspring were weaned at 28 days old and housed with same sex siblings. Upon weaning, litter size, sex, and weight were recorded. Offspring were kept on their respective diets until adults and released into the OPA enclosures. This duration of rofecoxib exposure maximized the ability of OPAs to detect health consequences as once released into the seminatural enclosures, all animals were fed the control diet because we are currently unable to keep animals on their respective diets while they are free ranging during OPAs. Switching the rofecoxib-exposed animals to the control diet was the most conservative approach of detecting fitness impacts as cumulative damage of the pharmaceutical prior to OPAs.

5.3.3 OPA enclosures. Enclosures have previously been described in (Ruff et al. 2013). Briefly, there are 11 independent enclosures approximately 30 m^2. Each enclosure is divided into six territories by hardwire mesh that is easily climbed but adds a component of spatial complexity. Optimal territories (n = 4) consisted of a large storage bin with multiple dark nesting sites and direct access to food. Suboptimal territories (n = 2) consisted of light-exposed nesting sites. All territories contained *ad libitum* access to food and water and were kept on a 12:12 hr light:dark cycle.

Five independent OPA populations were established and four were maintained for 28 weeks. One population was terminated at 11 weeks due to 100% control male mortality and because the enclosure was needed for another experiment. Populations consisted of eight to 10 males and 12-18 females for a total of 116 animals (42 male, 74 female). Half of each sex were from the rofecoxib-exposed treatment while the remaining half served as controls. The population structure allowed for direct competition between treatment and control individuals. Enclosure space and population size created a population density reported within the range observed in the wild (Sage 1981).

Upon enclosure release, male mice were on average 16.31 (SD ± 5.24) weeks old and females were 15.34 (SD ± 5.31) weeks old. To allow males to establish territories and prevent incidental matings, males from both treatments were released into the enclosures with nonexperimental females. One week later, nonexperimental females were removed and replace with experimental females from both treatments. Unrelated individuals at the cousin level or above was ideal for OPA populations. Three populations consisted of one set of brothers but no males were related to females. Four populations

consisted of one or two sister pairs, four populations consisted of sister triplets, and one population consisted of sister quadruplets. Relatedness was balanced between treatments.

5.3.4 Reproductive success. Founder reproductive success was determined by genetically analyzing offspring born in the enclosures. Offspring were first removed at week eight, then during five-week intervals, referred to as pup sweeps. This time interval, prevented offspring from reaching sexual maturity and from confounding the reproductive success data. During pup sweeps, offspring were removed, sacrificed, and a tissue sample was collected for genetic analyses. A total of 1,138 samples were collected with an average of 227.60 ± 104.50 (M ± S.D.) offspring per population.

A population-level approach was used to determine reproductive success in three of the five populations described in (Meagher et al. 2000). Briefly, founder individuals from each treatment were selected upon nonoverlapping sex-specific allelic variants, females upon the mitochondrial genome and males on the Y-chromosome. To control for confounding effects, such as segregating genes linked with the markers, reciprocal markers were assigned across populations. Mitochondrial genotypes were assessed in 860 samples (three of five populations) and obtained for 100% of offspring. Of the 1,138 offspring, 570 Y-chromosome genotypes were obtained (from all five populations), suggesting that 100% of all males were typed if the sex ratio was 1:1.

Reproductive success was determined by multiple microsatellite loci in one population. One population was left ungenotyped as the results of this population would not have influenced the overall trends of female reproductive success results to meet our predictions. Microsatellite data were converted to population levels readouts and combined with the mitochondrial and Y-chromosomal data for analysis. Between six and

11 autosomal microsatellite loci were amplified, scored, and analyzed in a stepwise fashion. Loci used were: d1mit251, d1mit449, d3mit22, d3mit312, d3mit333, d6mit138, d9mit232, d9mit251, d12mit277, d14mit128, and d19mit110. Primer sequences were obtained from the Mouse Genome Informatics website, The Jackson Laboratory, Bar Harbor Maine (http://www.informatics.jax.org/ accessed March 2014). Fluorescently tagged primers were used in PCR reactions (CY-5 or CY-3). Tagged PCR products were run on 14" x 17", 6.25% denaturing acrylamide gels at 40 W for three to seven hours (locus dependent). Gels were imaged on a Typhoon Scanner 8600 and ImageQuant software (Amersham Biosciences, Piscataway, NJ).

Parentage was assigned by using Cervus 3.0 (Kalinowski et al. 2007), a program that uses a likelihood based statistical approach. Allele frequencies were calculated using the genotypes of all candidate mothers and fathers and all offspring within the population. Simulations were run 10,000 cycles with an error rate of 1% to derive a delta score. Assigned parents were accepted when the trio confidence of mother, father, and offspring was 95%. With this rule, 86% (49/57) of the population was genotyped.

5.3.5 Male competitive ability. Prior to OPA release, all individuals received a PIT (passive integrated transponder) tag (TX1400ST, BioMark, Boise, ID) and a unique ear punch for identification purposes. Two sets of PIT antenna and readers (FS2001F-ISO, BioMark, Boise, ID) were used in this experiment and were rotated twice per week among concurrent populations. PIT tag antennas were placed above each feeding station within the six territories of an enclosure. All PIT tag data were downloaded to a computer containing data logging software (Minimon, Culver City, CA). A male was considered territorial occupant (dominant) in a territory if $> 80\%$ of the total reads belonged to him

at a particular location. PIT tag data were collected on female mice; however, they were not analyzed due to the lack of information on female dominance behavior.

5.3.6 Survivorship. Survivorship was assessed by daily noninvasive health checks and extensive enclosure checks during pup sweeps. Extensive checks were not performed at a higher frequency as to not disrupt territoriality formation, which increases infanticidal behavior. Research personnel entered enclosures only to freshen wateres, fill feeders, rotate PIT tag readers, remove deceased individuals, and to conduct pup sweeps. Deceased founders were identified by PIT tag IDs. The date of death was estimated upon the condition of the corpse. Individuals that had died long before research personnel discovered it were given a death date half way between the date it was found and the last date it was read by PIT tag readers.

5.4 Statistical Analyses.

Wean weight of offspring was analyzed with a linear mixed-effects model (LMM). This model assessed the effects of treatment, time, and the interaction of time and treatment. The model intercept was set to litter one. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. Sexes were analyzed separately. A normal distribution was assumed because weight data are continuous. Sample sizes used to assess weight differences include: 21 daughters from seven rofecoxib-exposed breeding pairs, 75 daughters from 16 control breeding pairs, and 26 sons from eight rofecoxib-exposed breeding pairs compared to 71 sons from 17 control breeding pairs.

Litter size data are discrete counts and therefore were analyzed with a generalized linear mixed model (GLMM) with a Poisson distribution and logarithmic link. This model assessed the effects of treatment, time, and the interaction of time and treatment. The model intercept was set to litter one. Treatment, parity, and their interaction were treated as fixed effects while cage was modeled as a random effect with a random intercept generated for each. There were 21 rofecoxib-exposed litters and 44 control litters for a total of 65 observations from 30 cages in this analysis. Reported SEMs are asymmetric because values have been back-transformed from logarithmic data.

Body weight was analyzed with a LMM. This model assessed the effects of treatment, time, and the interaction of time and treatment on the 116 population founders (females = 72, males = 42). Sexes were analyzed separately. A normal distribution was assumed because weight data are continuous. Treatment, time, and their interaction were modeled as fixed effects and individual and population were modeled as random effects with random slopes and intercepts generated for each. The intercept was set at week zero, as this was when founders were release into the enclosures and at which collected of weight data from OPAs began. Collection of founder weight data continued on surviving individuals at each pup sweep. There were a total of 337 female observations and 142 male observations collected throughout the experiment.

Reproductive outputs were in terms of total offspring and thus are discrete data. These data were analyzed with a GLMM with a Poisson distribution and logarithmic link. The model assessed the effects of treatment, time, and the interaction of treatment and time on population-level reproduction. These effects were set as fixed effects in the model and population was set as a random effect with random intercept calculated for

each. The intercept was set at week eight, as that was when the first collection period or pup sweep occurred. Male reproductive output for each treatment was measured five times over the course of the 28-week study in each of the five independent populations (except for one population that was only measured two times) for a total of 44 observations. Female reproductive output for each treatment was measured five times over the course of the 28-week study in four independent populations (except for one population that was only measured two times) for a total of 34 observations. Female reproductive output was analyzed in terms of total offspring and male reproductive success was analyzed in terms of male offspring. Reported SEMs are asymmetric because values have been back-transformed from logarithmic data.

To assess the probability of territorial ownership, a GLMM was used. As a territory can be defended or not, a binomial distribution was used with a logit link. There were six territories within a population and were either occupied by rofecoxib-exposed males, control males, or unoccupied. A total of 104 observations were collected and analyzed throughout the study. The model assessed the effect of treatment, time, and their interaction. These effects were set as fixed effects and population was set as a random effect with a random intercept generated for each. The model intercept was set to week three as that was when data existed for each population.

Survival of founders was analyzed by a multivariate Cox proportional hazard model. Impacts of treatment and population were examined in the model. In the male analyses, the interaction of treatment and population was also assessed. This was not possible for females because of low mortality ($n = 2$). Sexes were analyzed separately due to differences in mortality rates. Individuals that survived the length of the study or

that were intentionally removed from the study were censored. A total of 74 females were analyzed; two events and 72 censorings; and 42 males were analyzed; 17 events and 25 censorings.

Proportion hazard models were conducted in JMP 9.0.3 (SAS institute Inc., Cary NC). All mixed models were conducted in R 3.0.2 using either *glmer* or *lmer* functions within the lme4 library (Bates et al. 2014, R Development Core Team 2013). P values were calculated for LMM with the Swatterthwaite approximation under the lmerTest function (Kuznetsova et al. 2013). For all mixed models, several candidate models were fit to the data. These models varied in terms of random effects that estimated both intercept and/or slope. For each analysis, the model that explained some of the variance and with the lowest Akaike information criterion (AIC) score was selected and reported. Neither the significance of a fixed effects nor the magnitude of the significance varied between models.

5.5 Results

5.5.1 Breeding cage. No difference in wean weight was detected due to rofecoxib exposure (LMM; female offspring, $t = -0.80$, $p = 0.42$; male offspring, $t = -1.28$, $p = 0.20$). At the model intercept (litter one), rofecoxib-exposed female offspring weighed on average of 10.60 g (SEM ± 0.66) while controls weighed an average of 11.13 g (± 0.33). No effect of time (LMM; $t = -0.45$, $p = 0.65$) or time by treatment (LMM; $t = 0.76$, $p = 0.45$) was detected. Rofecoxib-exposed male offspring weighed on average 11.05 g (± 0.81) and controls weighed on average 12.09 g (± 0.37). There was a significant increase in male weight over time (LMM; $t = 2.85$, $p < 0.01$), but this trend was consistent

between treatments (LMM; $t = 0.04$, $p = 0.97$). For a complete readout of mixed model results for breeding cage data, see Table 5.1.

A trend was detected where rofecoxib-breeders produced smaller litters (GLMM; $z = -1.70$, $p = 0.09$; Figure 5.1). In first litters (model intercept), rofecoxib-exposed breeders produced an average of 3.09 pups (SEM +0.69, −0.57), while control breeders produced 4.36 pups (+0.48, −0.43). No effect of time (GLMM; $z = 0.30$, $p = 0.77$) or time by diet (GLMM; $z = 1.08$, $p = 0.28$) was detected, suggesting the trend of reduced litter size was consistent over time.

5.5.2 OPA measurements. Rofecoxib-exposure did not affect female body weight at the model intercept (week zero of LMM; $t = 0.21$, $p = 0.83$; Figure 5.2A). Rofecoxib-exposed females weighed on average 19.68 g (SEM ± 1.02) and controls weighed on average 19.47 g (± 0.97). Females gained weight over time, due to pregnancy (LMM; $t = 12.02$, $p < 0.0001$), and this effect was consistent between treatments (LMM; $t = -0.81$, $p = 0.42$). Rofecoxib-exposed males showed a marginally significant trend of reduced weight at the model intercept (week zero) when compared with controls (LMM; $t = -1.78$, $p = 0.08$; Figure 5.2B). Rofecoxib-exposed males weighed on average 20.13 g (± 0.73) and controls weighed on average 21.43 g (± 0.51). Males from both rofecoxib and control treatments gained weight over time (LMM; $t = 4.54$, $p < 0.0001$) and a treatment by time effect was detected (LMM; $t = 1.983$, $p < 0.05$), suggesting that rofecoxib-exposed males gained weight over time than did controls. For a complete readout of mixed model results for OPA weight measures data, see Table 5.2.

Rofecoxib-exposure increased female reproductive success by 40% relative to controls. At model intercept (week eight), rofecoxib-exposed females had significantly

more offspring than controls (GLMM; $z = 3.89$, $p < 0.0001$; Figure 5.3A) with a mean of 28.76 offspring per population (SEM +3.44, −3.07). Control females had a mean of 18.53 offspring per population (+4.24, −3.44). No effect of time (GLMM; $z = 0.39$, $p = 0.70$) or time by diet interaction (GLMM; $z = 1.20$, $p = 0.23$) was detected; suggesting rofecoxib-exposed females had more offspring throughout the duration of the study. For a complete readout of mixed model results for OPA reproduction and competitive ability, see Table 5.3.

Rofecoxib exposure did not impact male reproductive success (GLMM; $z = 0.19$, $p = 0.85$; Figure 5.3B). At model intercept (week eight), rofecoxib males sired on averaged 10.31 male offspring (SEM +1.87, −1.41) per population. Control males had an average of 10.03 male offspring (+1.96, −1.64) per population. No effect of time was detected (GLMM; $z = 1.38$, $p = 0.17$). A trend was detected in which rofecoxib-exposed males produced fewer male offspring over time (GLMM, $z = -0.05$, $p = 0.09$).

Male competitive ability was not impacted by treatment. At week three (model intercept), control males occupied 30% of territories, rofecoxib-exposed males occupied 30%, leaving 40% of territories unoccupied (GLMM; $z = 0.08$, $p = 0.94$; Figure 5.4). The percent of undefended territories is not unusual because 2/6 (or 33%) of the territories are suboptimal and often difficult to defend. There was a marginally significant increase in the number of territories being occupied over time (GLMM; $z = 1.72$, $p = 0.09$); but no time by diet interaction occurred (GLMM; $z = -0.27$, $p = 0.79$), suggesting that males from both treatments occupied more territories over time.

A trend was detected in which control female suffered more mortality (PH; $\chi^2 = 2.93$, $p = 0.09$; Figure 5.5A). However, this is based upon only two mortalities from 74

females. Mortality rate did not differ in replicate populations (PH; $\chi^2 = 6.15$, $p = 0.19$). No significant differences were detected in male mortality between treatments (PH; $\chi^2 = 0.04$, $p = 0.83$; Figure 5.5B). Mortality rate did not differ in replicate populations (PH; $\chi^2 = 4.85$, $p = 0.30$), nor was there a difference in the effect of treatment among populations (PH; $\chi^2 = 2.92$, $p = 0.57$).

5.6 Discussion

Rofecoxib-exposure did not affect survivorship, weight, male competitive ability, or male reproductive success within OPAs. Results from this experiment are conservative as rofecoxib exposure ceased when animals were released into enclosures and were all fed the control diet. One possible explanation for our observed results is that the drug did not cause lasting deleterious effects and after exposure ceased, exposed individuals rebound to average fitness. A second explanation is perhaps animals in this experiment were not exposed to rofecoxib long enough to induce fitness effects. Animals were exposed to rofecoxib during gestation and to ~15-16 weeks of age whereas cardiac adversity was not detected in humans until after taking the prescription for >18 months. Another possible explanation is that although cardiac events occurred in 27,000 people, an estimated 80 million people took the drug (Jüni et al. 2004), suggesting < 0.4% of people who took the drug had a cardiac event. Due to such a small percentage of cardiac adversity in humans, this effect was likely to go undetected in five populations of mice with 116 individuals.

Rofecoxib-exposed litters tended to be smaller when born in cages. This result is in line with previous research on NSAIDs that have shown these drugs can cause

transient infertility (Østensen et al. 2006), as prostaglandin synthesis is important for normal ovulation (Sugimoto 1997). Prostaglandin synthesis disruption was also suggested to be the cause of increased embryo loss in rofecoxib-exposed rats during preclinical studies (Merck 2002).

Although rofecoxib-exposed litters tended to be smaller in cages, when animals were in OPAs, rofecoxib-exposed females had 40% more offspring when compared to controls. One factor that may be affecting these results is that there were fewer populations than what is typical in OPA assessments. Female reproductive success was analyzed in four populations up to week 13 and only in three populations for the remaining portion of the study. Regardless of sample size, these contrasting results are interesting because the main variable that changed was the type of environment these animals were in and because increased reproduction was detected in a treatment where the immune system is targeted. Tradeoffs exist between pregnancy and the immune system and because rofecoxib suppresses the immune system, perhaps less energy was allocated towards immune function and more energy was allocated to pregnancy (Sheldon and Verhulst 1996). The fact that we detected a positive impact on female reproduction within seminatural environments is curious and warrants further investigation.

OPA assessment of rofecoxib is the second case in which no adverse health consequences were detected. In an unpublished study examining the health consequences of a nanomaterial (polyamidoamine dendrimers, PAMAM), no negative effects were detected. However, in four published OPA studies and in two pharmaceutical studies, at least one OPA endpoint (reproductive success, survivorship, or male competitive ability)

was negatively affected in at least one sex of treatment individuals (Figure 5.6) (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013, Chapter 2, Chapter 3). In all of these studies, OPAs revealed greater fitness consequences than in previous experimentation, thus providing support for the sensitivity of OPAs and because OPAs did not detect consequences in two substances does not negate the importance of implementing OPAs in toxicity assessment.

Although the detection of negative health consequences induced by rofecoxib exposure slipped the detection by OPAs, they were also missed by preclinical trials and we believe that OPAs would still be beneficial to be implemented within pharmaceutical development. In addition to the proximate-level animal studies, OPAs would be useful because they quantify fitness on an ultimate level that provides clear interpretation on overall health. This is achieved because of the nature of OPAs that challenge most physiological systems synergistically and simultaneously and are useful in revealing disease phenotypes that can then undergo further investigation for mechanistic cause. OPAs would be a useful tool to implement in pharmaceutical development and other areas of toxicity assessment.

5.7 Acknowledgments

I thank my PhD advisor, Wayne Potts for his support; James Ruff for statistical advise and valuable comments on this chapter; Michelle Bartlett, Stephanie Eddy, Tessa Galland, David Kircher, Kirstie Kandaris, Nicole Liu, Elyse Schwab, Mirtha Sosa, Tristan Underwood, Elizabeth Young, and Linda Morrison for enclosure maintenance, data collection, and genotyping; Ahmed Bwika and Jon Gale for animal care. This project

was funded by the University of Utah's Technology Commercialization Program and was partially conducted while Wayne Potts was supported by NSF grant DEB 09-18969 and NIH grant R01-GM109500 and I was supported by an NSF GK-12 Educational Outreach Fellowship (DGE 08-41233).

5.8 References

Bates, D., M. Maechler, B. Bolker, S. Walker, R. Christensen, and H. Singmann. 2014. lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-5.

Bombardier, C., L. Laine, A. Reicin, D. Shapiro, R. Burgos-Vargas, B. Davis, R. Day, M. Ferraz, C. Hawkey, M. Hochberg, T. Kvien, and T. Schnitzer. 2000. Comparion of Upper Gastrointestinal Toxicity of rofecoxib and Naproxen in Patients with Rheumatoid Arthritis. The New England journal of medicine 343:1520–8.

Carroll, L. S., S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. 2004. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. Evolution 58:1318–28.

Claycomb, R. J., S. J. Hewett, and J. A. Hewett. 2011. Prophylactic, prandial rofecoxib treatment lacks efficacy against acute PTZ-induced seizure generation and kindling acquisition. Epilepsia 52:273–83.

Cunningham, C. B., J. S. Ruff, K. Chase, W. K. Potts, and D. R. Carrier. 2013. Competitive ability in male house mice (Mus musculus): genetic influences. Behavior genetics 43:151–60.

Drazen, J. M. 2005. COX-2 Inhibitors — A Lesson in Unexpected Problems. The New England journal of medicine 352:1131–2.

Horton, R. 2004. Vioxx, the implosion of Merck, and aftershocks at the FDA. Lancet 364:1995–6.

Ilmonen, P., D. J. Penn, K. Damjanovich, J. Clarke, D. Lamborn, L. Morrison, L. Ghotbi, and W. K. Potts. 2008. Experimental infection magnifies inbreeding depression in house mice. Journal of evolutionary biology 21:834–41.

Jüni, P., L. Nartey, S. Reichenbach, R. Sterchi, P. A. Dieppe, and M. Egger. 2004. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. Lancet 364:2021–9.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular ecology 16:1099–106.

Keane, W. 2004. Merck Announces Voluntary Worldwide Withdrawal of VIOXX. Accessed April 2014 http://www.pbm.va.gov/vacenterformedicationsafety/vioxx/DearHealthcareProfessional.pdf.

Krumholz, H. M., J. S. Ross, A. H. Presler, and D. S. Egilman. 2007. What have we learnt from Vioxx? British Medical Journal 334:120–3.

Kuznetsova, A., P. Brockhoff, R. Christensen. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-3.

Lipsky, M. S., and L. K. Sharp. 2001. From idea to market: the drug approval process. The Journal of the American Board of Family Practice 14:362–7.

Loftus, P. 2010. Merck Books $950 Million Vioxx Charge. Accessed April 2014 http://online.wsj.com/article/SB10001424052702304155604575581923223447954.html.

Meagher, S., D. J. Penn, and W. K. Potts. 2000. Male-male competition magnifies inbreeding depression in wild house mice. Proceedings of the National Academy of Sciences of the United States of America 97:3324–9.

Merck. 2002. VIOXX® (rofecoxib tablets and oral suspension). Accessed April 2014. http:// www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4090B1_14_O-FDA-Tab-I-1.htm.

Miller, H. 2012. How is the FDA really doing? Genetic Engineering and Biotechnology News 32:6–8.

Nelson, A. C., J. W. Cauceglia, S. D. Merkley, N. A. Youngson, A. J. Oler, R. J. Nelson, B. R. Cairns, E. Whitelaw, and W. K. Potts. 2013. Reintroducing domesticated wild mice to sociality induces adaptive transgenerational effects on MUP expression. Proceedings of the National Academy of Sciences 110:19848–53.

Østensen, M., M. Khamashta, M. Lockshin, A. Parke, A. Brucato, H. Carp, A. Doria, R. Rai, P. Meroni, I. Cetin, R. Derksen, W. Branch, M. Motta, C. Gordon, G. Ruiz-Irastorza, A. Spinillo, D. Friedman, R. Cimaz, A. Czeizel, J. C. Piette, R. Cervera, R. a Levy, M. Clementi, S. De Carolis, M. Petri, Y. Shoenfeld, D. Faden, G. Valesini, and A. Tincani. 2006. Anti-inflammatory and immunosuppressive drugs and reproduction. Arthritis research & therapy 8:209.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reagan-Shaw, S., M. Nihal, and N. Ahmad. 2008. Dose translation from animal to human studies revisited. The Federation of American Societies for Experimental Biology Journal 22:659–61.

Ricciotti, E., and G. a FitzGerald. 2011. Prostaglandins and inflammation. Arteriosclerosis, thrombosis, and vascular biology 31:986–1000.

Ruff, J. S., A. K. Suchy, S. a Hugentobler, M. M. Sosa, B. L. Schwartz, L. C. Morrison, S. H. Gieng, M. K. Shigenaga, and W. K. Potts. 2013. Human-relevant levels of added sugar consumption increase female mortality and lower male fitness in mice. Nature communications 4:DOI 10:1038/ncomms3245.

Sage, R. 1981. The Mouse in Biomedical Research. Pages 40–90 (H. Foster, J. Small, and J. Fox, Eds.). Academic Press, New York.

Schuster, D., C. Laggner, and T. Langer. 2005. Why Drugs Fail - A Study on Side Effects in New Chemical Entities. Current Pharmaceutical Design 11:3545–59.

Sheldon, B. C., and S. Verhulst. 1996. Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. Trends in Ecology and Evolution 5347:317–21.

Sugimoto, Y. 1997. Failure of Parturition in Mice Lacking the Prostaglandin F Receptor. Science 277:681–3.

The Jackson Laboratory. 2013. Accessed March 2014. http://www.informatics.jax.org.

Table 5.1. Mixed model results for litter size and wean weight from breeding cages.

Litter Size	GLMM with Poisson distribution and logarithmic link (groups = 30, observations = 65)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.0173	0.1316		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Intercept (Litter 1)	1.4721	0.1047	14.0590	<0.0001***
Treatment (Rofecoxib)	-0.3432	0.2020	-1.6990	0.0894
Parity	0.0267	0.0903	0.2950	0.7678
Treatment (Rofecoxib)×Parity	0.2085	0.1929	1.0800	0.2800
Female Wean Weight	LMM (groups = 23, observations = 96)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	0.8188	0.9049		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(>|t |)*
Intercept (Litter 1)	11.1311	0.3242	34.3330	<0.0001***
Treatment (Rofecoxib)	-0.5296	0.6608	-0.8010	0.4270
Parity	-0.0937	0.2064	-0.4540	0.6510
Treatment (Rofecoxib)×Parity	0.3591	0.4697	0.7640	0.4470
Male Wean Weight	LMM (groups = 25, observations = 97)			
Random effects	*Variance*	*Std. Dev.*		
Cage (Intercept)	1.0400	1.0200		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(>|t |)*
Intercept (Litter 1)	12.0867	0.3715	32.5300	<0.0001***
Treatment (Rofecoxib)	-1.0360	0.8101	-1.2800	0.2048
Parity	0.6496	0.2277	2.8500	0.0054**
Treatment (Rofecoxib)×Parity	0.0250	0.5889	0.0400	0.9663

** Indicates a p value < 0.01, ***< 0.001.

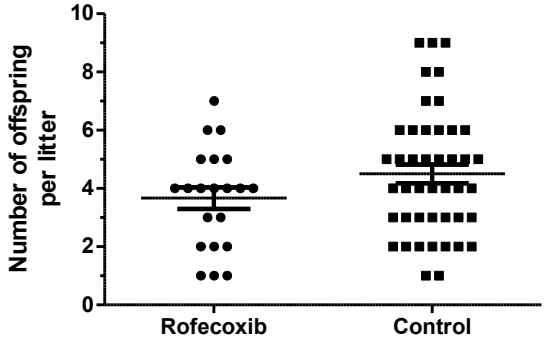

Figure 5.1. Litter size at weaning age from all litters produced by rofecoxib-exposed and control breeders. A trend was detected where rofecoxib-exposed breeders produced smaller litters [$n = 30$, *observations* = 65 (GLMM; $z = -1.70$, $p = 0.09$)]. Lines represent mean and standard error.

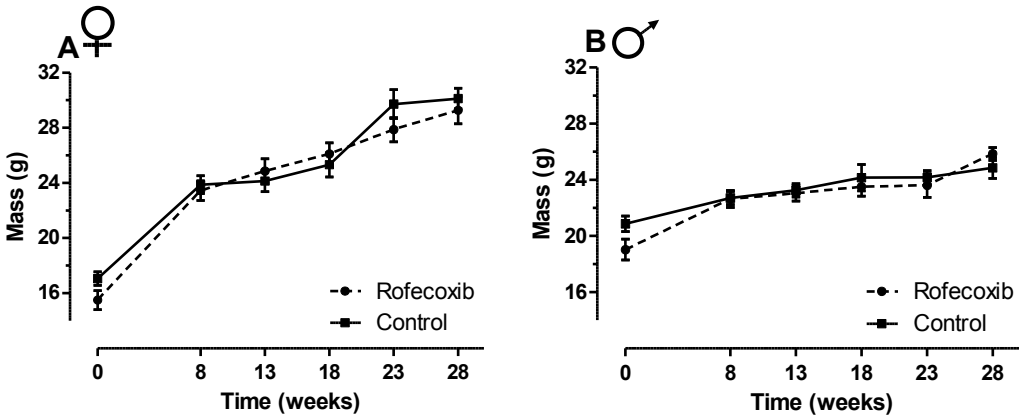

Figure 5.2. Body weight of rofecoxib-exposed and control founders within OPAs over time. A) No differences in body weight between treatments were detected in females at model intercept [*n* = 74, *observations* = 337 (week zero of LMM; t = 0.21, p = 0.83)]. Both rofecoxib-exposed and control female gained weight over time (LMM; t = 12.02, p < 0.0001). B) A trend was detected where rofecoxib-exposed males weighed less than controls at the model intercept [*n* = 39, *observations* = 142 (week 0 of LMM; t = −1.78, p = 0.08)]. Males gained weight over time (LMM; t = 4.54, p < 0.0001) and rofecoxib-exposed males gained more weight over time than controls (LMM: t = 1.98, p < 0.05). Lines connect means and error bars represent standard error.

Table 5.2. Linear mixed model results for body weight over time within OPAs.

Female Body Weight	LMM (groups = 74, observations = 337)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	9.2891	3.0478		
Population (Intercept)	1.9916	1.4112		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(>|t |)*
Intercept (Week 0)	19.4696	0.9668	20.1380	<0.0001***
Treatment (Rofecoxib)	0.2144	1.0208	0.2100	0.8340
Time	0.3916	0.0326	12.0220	<0.0001***
Treatment (Rofecoxib)×Time	-0.0357	0.0442	-0.8080	0.4200
Male Body Weight	LMM (groups = 39, observations = 142)			
Random effects	*Variance*	*Std. Dev.*		
Individual (Intercept)	2.3450	1.5314		
Fixed effects	*Estimate*	*Std. Error*	*t value*	*Pr(>|t |)*
Intercept (Week 0)	21.4308	0.5111	41.9270	<0.0001***
Treatment (Rofecoxib)	-1.3016	0.7323	-1.7770	0.0793
Time	0.1116	0.0246	4.5420	<0.0001***
Treatment (Rofecoxib)×Time	0.0715	0.0361	1.9830	0.0495*

* Indicates a p value < 0.05, *** < 0.001.

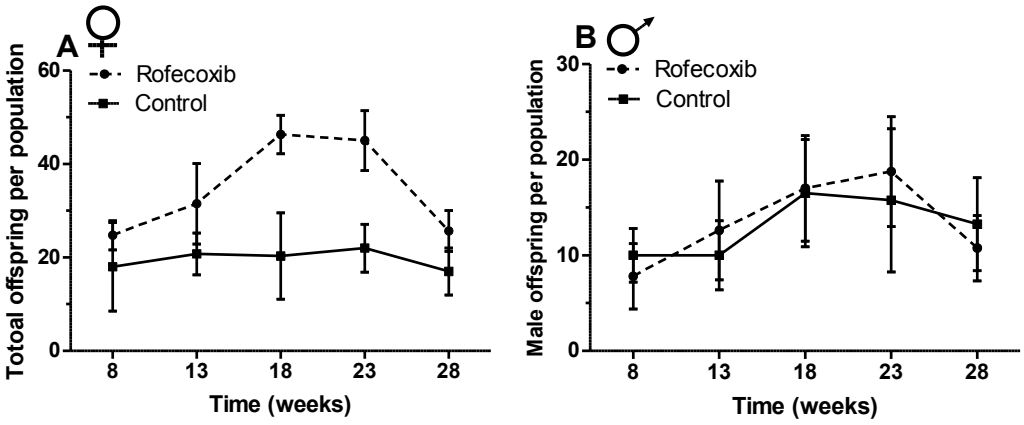

Figure 5.3. Reproductive success of rofecoxib-exposed and control animals in OPAs. A) Rofecoxib-exposed females had 40% more offspring than controls [$n = 4$, *observations* = 34 (GLMM; $z = 3.89$, $p < 0.0001$)] and this effect was consistent throughout the study. B) No difference in male reproductive success was detected between treatments [$n = 5$, *observations* = 44 (GLMM; $z = 0.19$, $p = 0.85$)]. A trend was detected in which rofecoxib males had more male offspring over time (GLMM; $z = -0.05$, $p = 0.09$). Female reproduction is in terms of total offspring as mitochondrial and multiple microsatellite markers that were used. Male reproduction is in terms of male offspring due to using Y-chromosome markers. Lines connect means of the populations at each time point for each sex and error bars represent standard error.

Table 5.3. Generalized linear mixed model results for reproduction and male competitive ability over time within OPAs.

Female Reproduction	GLMM with Poisson distribution and logarithmic link (groups = 4, observations = 34)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.1755	0.4189		
Population (Slope)	0.0008	0.0283		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Intercept (Week 8)	2.8937	0.2283	12.6760	<0.0001***
Treatment (Rofecoxib)	0.4393	0.1129	3.8910	<0.0001***
Time	0.0063	0.1616	0.3900	0.6960
Treatment (Rofecoxib)×Time	0.0116	0.0097	1.2040	0.2290
Male Reproduction	GLMM with Poisson distribution and logarithmic link (groups = 5, observations = 44)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept)	0.1036	0.3219		
Population (Slope)	0.0002	0.0150		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Intercept (Week 8)	2.3056	0.1786	12.9100	<0.0001***
Treatment (Rofecoxib)	0.0273	0.1468	0.1860	0.8530
Time	0.0156	0.0113	1.3830	0.1670
Treatment (Rofecoxib)×Time	-0.0006	0.0117	-0.0520	0.0959
Male Competitive Ability	GLMM with binomial distribution and logit link (groups = 5, observations = 104)			
Random effects	*Variance*	*Std. Dev.*		
Population (Intercept week 3)	0.1287	0.3587		
Fixed effects	*Estimate*	*Std. Error*	*z value*	*Pr(>|z|)*
Intercept (Week 3)	-0.8640	0.2525	-3.4220	0.0006***
Treatment (Rofecoxib)	0.0208	0.2767	0.0750	0.9401
Time	0.0272	0.0158	1.7210	0.0853
Treatment (Rofecoxib)×Time	-0.0059	0.0221	-0.2670	0.7895

*** Indicates a p value < 0.001.

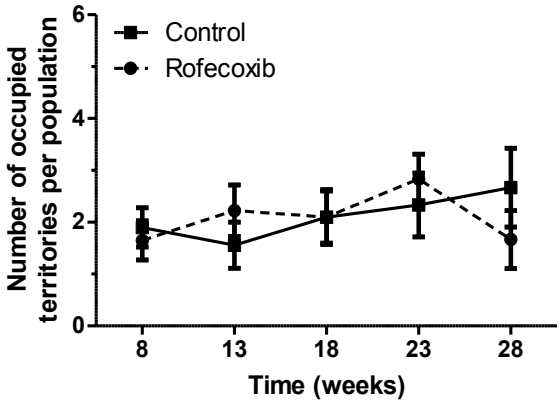

Figure 5.4. Male competitive ability between rofecoxib-exposed males and controls. Rofecoxib-exposed males and control males occupied the equal percentages of territories, 30% each [n = 5, *observations* = 104 (GLMM; z = 0.08, p = 0.94)]. This effect was consistent throughout the study. A male was considered dominant, or a territorial occupant, when > 80% of his reads were at a particular location. Points represent the number of territories occupied by males from each treatment over a multiday reading frame. Lines represent the best fit of the data. Paired observations (one rofecoxib point for every control) occur at each reader session. When data points from multiple populations overlapped, x-axis scatter was created for visual clarity.

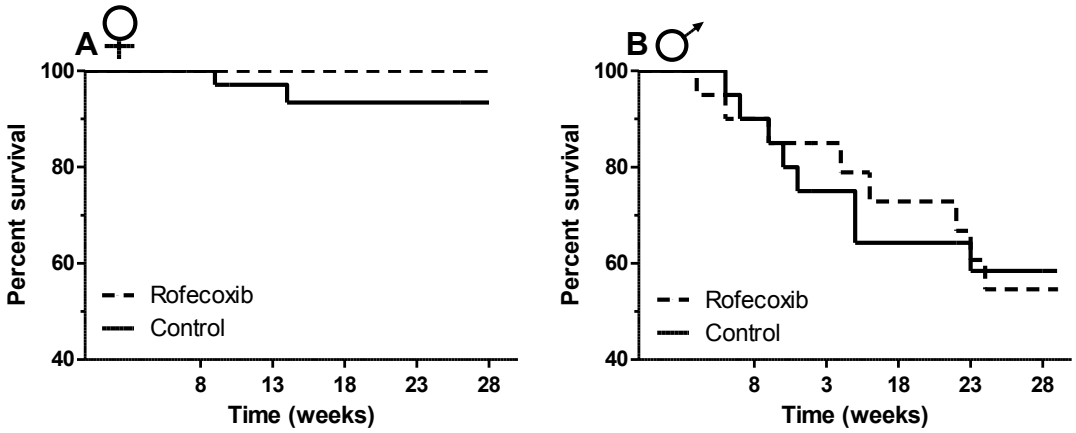

Figure 5.5. Survivorship of rofecoxib-exposed animals compared to controls in OPAs. A) A trend was detected where control females suffered higher mortality than rofecoxib-exposed females [n = 74 (PH; $\chi^2 = 2.93$, p = 0.09)]; however, this is based upon two mortalities. B) No differences in male mortality was detected between treatments [n = 42 (PH $\chi^2 = 0.04$; p = 0.83)].

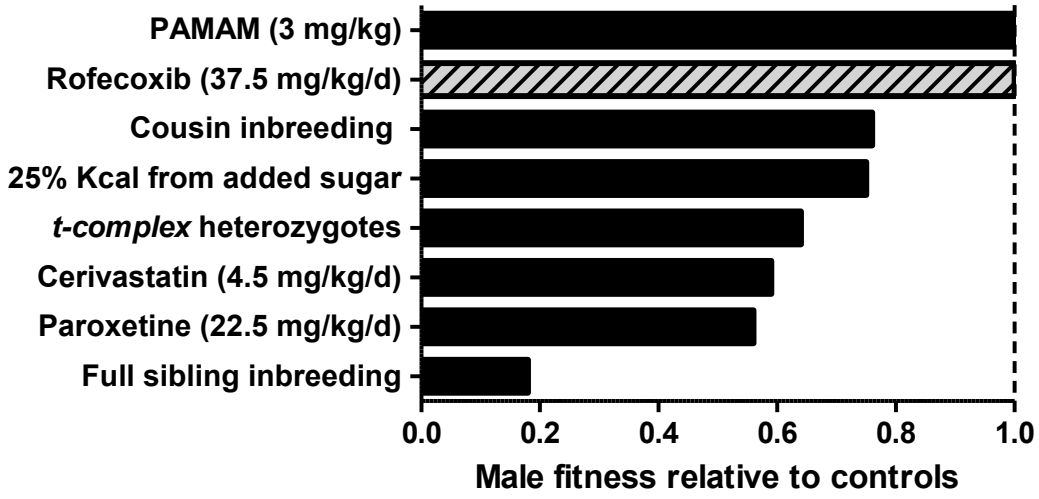

Figure 5.6. Comparison of male reproductive success relative to control counterpart from previous OPA experiments. Rofecoxib did not affect male fitness when compared with controls. Rofecoxib males had higher fitness when compared with both cousin level and sibling level inbreeding, 25% high fructose corn syrup consumption, a selfish gene, and the exposure of cerivastatin and paroxetine (Meagher et al. 2000, Carroll et al. 2004, Ilmonen et al. 2008, Ruff et al. 2013, Chapter 2, Chapter 3).